

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 23, 2001, 16:14:49 ; Search time 18.65 Seconds
(without alignments)
45,976 Million cell updates/sec

Title: US-09-529-691-1
Perfect score: 92
Sequence: 1 GVKGDKGNPGWGPAP 15

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 390729 seqs, 57163235 residues
Total number of hits satisfying chosen parameters: 390729

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 08
Maximum Match 100%
Listing first 45 summaries

Database :
1: /SID56/gcgdata/geneseq/AA1980.DAT.*
2: /SID56/gcgdata/geneseq/AA1981.DAT.*
3: /SID56/gcgdata/geneseq/AA1982.DAT.*
4: /SID56/gcgdata/geneseq/AA1983.DAT.*
5: /SID56/gcgdata/geneseq/AA1984.DAT.*
6: /SID56/gcgdata/geneseq/AA1985.DAT.*
7: /SID56/gcgdata/geneseq/AA1986.DAT.*
8: /SID56/gcgdata/geneseq/AA1987.DAT.*
9: /SID56/gcgdata/geneseq/AA1988.DAT.*
10: /SID56/gcgdata/geneseq/AA1989.DAT.*
11: /SID56/gcgdata/geneseq/AA1990.DAT.*
12: /SID56/gcgdata/geneseq/AA1991.DAT.*
13: /SID56/gcgdata/geneseq/AA1992.DAT.*
14: /SID56/gcgdata/geneseq/AA1993.DAT.*
15: /SID56/gcgdata/geneseq/AA1994.DAT.*
16: /SID56/gcgdata/geneseq/AA1995.DAT.*
17: /SID56/gcgdata/geneseq/AA1996.DAT.*
18: /SID56/gcgdata/geneseq/AA1997.DAT.*
19: /SID56/gcgdata/geneseq/AA1998.DAT.*
20: /SID56/gcgdata/geneseq/AA1999.DAT.*
21: /SID56/gcgdata/geneseq/AA2000.DAT.*
22: /SID56/gcgdata/geneseq/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	92	100.0	15	12	R12973
2	92	100.0	15	12	R83639
3	92	100.0	15	16	R82922
4	92	100.0	15	17	W01139
5	92	100.0	15	18	W13022
6	92	100.0	15	19	W56114
7	92	100.0	15	20	Y04481
8	92	100.0	15	21	Y78574
9	92	100.0	24	19	W46044
10	92	100.0	27	18	W13025
11	92	100.0	39	19	W46046

12	92	100.0	39	19	W46043	Collagen sequence
13	92	100.0	43	18	W13024	Collagen alpha1(IV)
14	86	93.5	406	21	B58169	Lung cancer associ
15	69	75.0	772	13	R23873	Human alpha 5 (IV)
16	69	75.0	772	18	W09643	Human type IV coll
17	64	69.6	430	21	B58180	Lung cancer associ
18	61	66.3	532	19	W40114	Human alpha-1(IV)
19	60	65.2	1078	16	R71704	Collagen alpha 1 (
20	60	65.2	1078	21	Y96125	Collagen type III
21	60	65.2	1196	13	R28916	Type III procollag
22	59	64.1	471	16	R79163	Partial sequence o
23	59	64.1	471	20	Y44171	Bovine type IV col
24	59	64.1	471	21	Y56783	Bovine alpha3 type
25	58	63.0	1288	20	W92297	Mouse alpha-1 (XVI
26	58	63.0	1301	20	W92296	Human alpha-1 (XVI
27	58	63.0	1336	20	Y08694	Human collagen 18
28	57	62.0	259	21	Y66642	Membrane-bound pro
29	57	62.0	259	21	Y44488	ACRP30R2 variant p
30	57	62.0	260	21	Y91522	Human secreted pro
31	57	62.0	278	22	B49596	Human zacrpf. Hom
32	57	62.0	287	21	Y91666	Human secreted pro
33	56	60.9	306	22	B27230	Human EXMD-8 SEO
34	56	60.9	489	19	W39749	Mouse marcosR prot
35	56	60.9	518	17	W03561	Macrophage recepto
36	56	60.9	546	19	W40115	Human alpha-5(IV)
37	55	59.8	50	21	B43060	Human ORFX ORF2824
38	55	59.8	96	21	G00125	Human secreted pro
39	55	59.8	96	21	Y64662	Human 5' EST relat
40	55	59.8	171	18	W18781	Recombinant bovine
41	55	59.8	171	18	W13673	Recombinant bovine
42	55	59.8	245	21	B30234	Human complement C
43	55	59.8	245	21	B43063	Human ORFX ORF2827
44	55	59.8	245	22	B49594	Human Cgl c protel
45	55	59.8	245	22	B49600	Human Ctg c protel

ALIGNMENTS

RESULT	1	ALIGNMENTS
12973		
ID	R12973 standard: peptide; 15 AA.	
AC	R12973;	
XX		
DT	24-SEP-1991 (first entry)	
DE	IV-HI based on type IV collagen alpha 1 chain triple helical region.	
XX		
KW	malignant cell growth; melanoma cell motility; cellular adhesion.	
XX		
OS	Synthetic.	
PN	W09108755-A.	
PD	27-JUN-1991.	
XX		
PF	06-DEC-1990: 90MO-US07162.	
XX		
PR	14-DEC-1989: 89US-0450419.	
XX		
PA	(MINU) MINNESOTA UNIVERSITY.	
XX		
PI	Chelberg MK, Tillybery PE, McCarthy JB;	
XX		
DR	WPI; 1991-207864/28.	
XX		
PT	Polypeptide for promoting tumour cell adhesion, spreading and	
XX	motility - useful in chemo:diagnosis and chemotherapy of	
PS	malignant cell growth	
XX		
XX	Claim 1: Page 26; 37pp; English.	

CC This polypeptide (designated IV-H1) corresponds to type IV collagen
 CC residues 1263-1277 from the major triple helical region of the alpha
 CC 1 chain type of IV collagen. It, or a polypeptide comprising this
 CC sequence, can be used to promote cellular attachment to substrata
 CC or to inhibit the metastasis and invasion of tumour cells.

XX SQ Sequence 15 AA;

Query Match 100.0%; Score 92; DB 12; Length 15;
 Best Local Similarity 100.0%; Pred. No. 2.6e-07;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GVKGDKNPMPGAP 15
 |||||
 DB 1 gvkgdgmpgwpgap 15

RESULT 2

R83639 standard; peptide; 15 AA.

AC R83639;

XX 11-APR-1996 (first entry)

XX Collagen IV alpha-2 chain NC1 domain peptide IV-H1 (1263-1277).

XX Collagen; type IV; alpha-2 chain; NC1 domain; fibroblast; cell;

XX proliferation; suppression; epithelial; glial; epiretinal; eye;

XX membranes; contraction; migration; vitreoretinopathy; scarring;

XX vitreoretinal; glaucoma.

XX Synthetic.

XX WO9522979-A1.

XX 31-AUG-1995.

XX 28-FEB-1995; 95WO-0502478.

XX 27-FEB-1995; 95US-0203458.

XX 28-FEB-1994; 94US-0203458.

XX (MINU) UNIV MINNESOTA.

XX PA

XX PI

XX PI

XX PI

XX PI

XX PI

XX PI

XX PI

XX PI

XX PI

XX PI

XX PI

XX PI

XX PI

OY 1 GVKGDKNPMPGAP 15
 |||||
 DB 1 gvkgdgmpgwpgap 15

RESULT 3

R82922 standard; peptide; 15 AA.

AC R82922;

XX 20-DEC-1995 (first entry)

XX Non-RGD, non-YISGR cancer metastasis inhibitory peptide #16.

XX Cancer metastasis; adhesive peptide; core sequence; dextran; cancer;

XX water soluble polysaccharide; metastasis; wound; immunogenicity.

XX Synthetic.

XX JP07089999-A.

XX 04-APR-1995.

XX 17-SEP-1993; 93JP-0254779.

XX 17-SEP-1993; 93JP-0254779.

XX (JAPG) NIPPON ZEON KK.

XX WPI: 1995-167254/22.

XX Cancer metastasis inhibitory peptide derivs. - useful for inhibition

XX of cancer metastasis, healing of wounds and regulation of

XX immunogenicity.

XX Disclosure: Page 3; 6pp; Japanese.

XX The peptides R70472-90 and R82907-24 are peptide derivatives which

XX inhibit cancer metastasis. They are composed of an adhesive peptide with

XX a core sequence selected from: RGP (R70472-85), YIGSR (R70486-90) or

XX other sequence (R82907-24), linked to a water soluble polysaccharide,

XX CC preferably a water soluble dextran, at the C-terminus. The peptides are

XX CC useful in inhibiting cancer metastasis, healing wounds and the regulation

XX CC of immunogenicity.

XX SQ Sequence 15 AA;

Query Match 100.0%; Score 92; DB 16; Length 15;
 Best Local Similarity 100.0%; Pred. No. 2.6e-07;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GVKGDKNPMPGAP 15
 |||||
 DB 1 gvkgdgmpgwpgap 15

RESULT 4

W01139 standard; peptide; 15 AA.

AC W01139;

XX 18-DEC-1996 (first entry)

XX Peptide 10 for glia cell removal derived from collagen.

XX glia cell; neuron; analysis; behaviour; selective; removal.

XX W01139 standard; peptide; 15 AA.

AC W01139;

XX 18-DEC-1996 (first entry)

XX Peptide 10 for glia cell removal derived from collagen.

XX glia cell; neuron; analysis; behaviour; selective; removal.

XX W01139 standard; peptide; 15 AA.

AC W01139;

XX 18-DEC-1996 (first entry)

XX Peptide 10 for glia cell removal derived from collagen.

XX glia cell; neuron; analysis; behaviour; selective; removal.

XX W01139 standard; peptide; 15 AA.

AC W01139;

XX 18-DEC-1996 (first entry)

XX Peptide 10 for glia cell removal derived from collagen.

XX glia cell; neuron; analysis; behaviour; selective; removal.

XX W01139 standard; peptide; 15 AA.

AC W01139;

XX 18-DEC-1996 (first entry)

XX Peptide 10 for glia cell removal derived from collagen.

XX glia cell; neuron; analysis; behaviour; selective; removal.

XX W01139 standard; peptide; 15 AA.

AC W01139;

XX 18-DEC-1996 (first entry)

XX Peptide 10 for glia cell removal derived from collagen.

XX glia cell; neuron; analysis; behaviour; selective; removal.

XX W01139 standard; peptide; 15 AA.

AC W01139;

XX 18-DEC-1996 (first entry)

XX Peptide 10 for glia cell removal derived from collagen.

XX glia cell; neuron; analysis; behaviour; selective; removal.

XX W01139 standard; peptide; 15 AA.

AC W01139;

XX 18-DEC-1996 (first entry)

XX Peptide 10 for glia cell removal derived from collagen.

XX glia cell; neuron; analysis; behaviour; selective; removal.

FH Key Location/Qualifiers
 PT Modified-site 15 /note="Pro-NH2"
 FT
 XX JP08073495-A.
 XX
 PD 19-MAR-1996.
 XX
 PF 31-AUG-1994: 94JP-0232263.
 XX
 PR 31-AUG-1994: 94JP-0232263.
 XX
 PA (AGEN) AGENCY OF IND SCI & TECHNOLOGY.
 XX
 DR WPI: 1996-205531/21.
 XX
 PT Selective remover for gila cells from neuronal cell cultures -
 PT useful for studying behaviour of neurons in the absence of gila
 PT cells.
 XX
 PS Example 1: Page 4: 6pp; Japanese.
 XX
 CC W01135-40 are peptides, derived from laminin, fibronectin and collagen,
 CC which selectively remove gila cells in the presence of neurons. This
 CC enables analysis of behaviour of neurons in the absence of gila cells.
 XX
 SO Sequence 15 AA:

Query Match 100.0%; Score 92; DB 17; Length 15;
 Best Local Similarity 100.0%; Pred. No. 2.6e-07;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 1 GVKGDKNPGMPCAP 15
 DB 1 gvkgdksnpqpgap 15

RESULT 5
 ID W13022
 XX W13022 standard; peptide: 15 AA.
 AC W13022:
 XX
 DT 30-APR-1997 (first entry)
 XX
 DE Collagen alpha1(IV) chain residues 1263-1277.
 XX
 KW Collagen: cell; adhesion; migration; promotion; alpha1(IV) chain;
 KW triple helical polypeptide; solid phase synthesis; study;
 KW O-glycosidic bond; deuterium label; structure; biological activity.
 XX
 OS Homo sapiens.
 XX
 PN USS576419-A.
 XX
 PD 19-NOV-1996.
 XX
 PF 30-JUN-1993: 93US-0085633.
 XX
 PR 30-JUN-1993: 93US-0085633.
 PR 27-SEP-1995: 95US-0534342.
 XX
 PA (MINU) UNIV MINNESOTA.
 XX
 PI Fields GB;
 XX
 DR WPI: 1997-011309/01.
 XX
 PT Solid phase synthesis of triple-helical branched polypeptide -
 PT under O-glycosidic bond and deuterium label retaining conditions,
 PT which may contain collagen cell adhesion sequences, useful for
 PT studying structure and biological activity of collagen

Configuration

XX
 PS Claim 18: Column 25: 25pp; English.
 XX
 CC The present peptide is a collagen cell adhesion sequence (CCAS),
 CC known to promote adhesion and migration of various cells,
 CC comprising residues 1263-1277 of the collagen alpha1(IV) chain.
 CC The triple helical polypeptide 2 (THP-2), which includes the
 CC present CCAS (residues 25-39) in its triple helix including region,
 CC was prepared by solid phase synthesis under O-glycosidic bond and
 CC deuterium label retaining conditions. THP-2 can be used to study
 CC the structure and biological activity of collagen.
 XX
 SO Sequence 15 AA:

Query Match 100.0%; Score 92; DB 18; Length 15;
 Best Local Similarity 100.0%; Pred. No. 2.6e-07;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 1 GVKGDKNPGMPCAP 15
 DB 1 gvkgdksnpqpgap 15

RESULT 6
 ID W56114
 XX W56114 standard; peptide: 15 AA.
 AC W56114:
 XX
 DT 09-JUL-1998 (first entry)
 XX
 DE Peptide representing the alpha-1(IV)1263-1277 collagen sequence.
 XX
 KW Alpha-1(IV)1263-1277 collagen sequence; peptide-amphiphile complex;
 KW lipophilic portion; stable lipid film; bilayer membrane system;
 KW drug-targeting; melanoma cell.
 XX
 OS Synthetic.
 XX
 PN W09807752-A1.
 XX
 PD 26-FEB-1998.
 XX
 PF 27-AUG-1996: 96MO-US14145.
 XX
 PR 23-AUG-1996: 96US-0702254.
 XX
 PA (MINU) UNIV MINNESOTA.
 XX
 PI Fields GB, Tirrell MV;
 XX
 DR WPI: 1998-179051/16.
 XX
 PT Peptide-amphiphile complexes which are able to form micelles or
 PT vesicles - comprise a lipophilic portion and a peptide portion which
 PT has a secondary structure
 XX
 PS Claim 15: Page 23: 33pp; English.
 XX
 CC The present sequence represents the alpha-1(IV)1263-1277 collagen
 CC sequence. This sequence is used to construct a peptide-amphiphile
 CC complex to exemplify the invention. The peptide-amphiphile complex of
 CC the specification comprises a lipophilic portion and a peptide portion
 CC having a secondary structure. The peptide portion preferably has
 CC biological activity, e.g. cell recognition activity or enzymatic
 CC activity. The complex may be used for forming stable lipid films on
 CC substrates, or may be used in bilayer membrane systems. It may, in
 CC micellar or vesicular form, be used in preparation of drug-targeting
 CC systems against melanoma cells.
 XX
 SO Sequence 15 AA:

Query Match 100.0%; Score 92; DB 19; Length 15;
 Best Local Similarity 100.0%; Pred. No. 2.6e-07;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GVKDGKGNPMPGAP 15
 |||||
 DB 1 gvkdgkgnpmpgap 15

RESULT 7

ID Y04481 standard; peptide: 15 AA.

AC Y04481;

DT 05-JUL-1999 (first entry)

DE Human type IV collagen alpha 1 chain D form fragment D-IVH1.

KW Human; type IV collagen alpha 1 chain; D-IVH1; Inhibition: tumour;
 cell adhesion; major triple helical domain; collagenous region.

OS Homo sapiens.
 OS Synthetic.

Key Location/Qualifiers
 Misc-difference 1..15
 FT /note="D-form residues"

W09920300-A1.

PD 29-APR-1999.

PF 22-OCT-1998; 98MO-US22405.

PR 22-OCT-1997; 97US-0062716.

PR 22-OCT-1997; 97US-0062617.

PA (MINU) UNIV MINNESOTA.

PI Fields GB, McCarthy JB;

DR WPI: 1999-302644/25.

PT New polypeptide useful for inhibition of tumour cell adhesion to type
 IV collagen

PS Claim 4; Page 18; 31pp; English.

XX The present sequence represents an all D-form fragment of the continuous
 CC collagenous region of the major triple helical domain of the alpha 1
 CC chain of human type IV collagen, designated D-IVH1. The peptide D-IVH1
 CC can be used to inhibit tumour cell binding to type IV collagen, inhibit
 CC tumour cell invasion of basement membrane and inhibit tumour cell
 CC metastasis in vivo. The peptide can be used to treat melanoma cell
 CC metastasis, and spontaneous Lewis lung tumour metastasis. The peptide is
 CC also useful for targeting cytotoxic agents and drugs to tumour cells.
 XX
 SQ Sequence 15 AA;

Query Match 100.0%; Score 92; DB 20; Length 15;
 Best Local Similarity 100.0%; Pred. No. 2.6e-07;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GVKDGKGNPMPGAP 15
 |||||
 DB 1 gvkdgkgnpmpgap 15

RESULT 8

Y78574

ID Y78574 standard; peptide: 15 AA.

AC Y78574;

DT 05-MAY-2000 (first entry)

DE Type IV collagen derived peptide IV-HI.

KW Type IV collagen; intracellular fibroblast proliferation inhibitor;
 proliferative retinopathy; fibroblast migration; scarring; glaucoma.

OS Unidentified.

PN US6013628-A.

PD 11-JAN-2000.

PF 27-FEB-1995; 95US-0394748.

PR 28-FEB-1994; 94US-0203458.

PA (MINU) UNIV MINNESOTA.

PI Gregerson DS, Agarwal A, Wright MM, Murali S, Skubitz APN;
 Furcht LT, Balles M;

DR WPI: 2000-159882/14.

PT Treatment of glaucoma by ocular administration of a type IV collagen
 peptide that inhibits fibroblast proliferation

PS Example 1; Column 15; 44pp; English.

XX This sequence represents a type IV collagen peptide. Ocular
 CC administration of the peptide inhibits fibroblast proliferation and can
 CC be used in the treatment of glaucoma. The peptide can be used to treat
 CC proliferative vitreoretinopathy by ocular administration of the peptides
 CC of the invention, which suppress fibroblast-mediated collagen gel
 CC contraction. The peptides can be used for treating glaucoma in mammals,
 CC especially humans, by inhibiting intraocular fibroblast proliferation,
 CC fibroblast migration and scarring.

SQ Sequence 15 AA;

Query Match 100.0%; Score 92; DB 21; Length 15;
 Best Local Similarity 100.0%; Pred. No. 2.6e-07;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GVKDGKGNPMPGAP 15
 |||||
 DB 1 gvkdgkgnpmpgap 15

RESULT 9

ID W46044 standard; peptide: 24 AA.

AC W46044;

DT 23-DEC-1998 (first entry)

DE Collagen sequence synthesised as triple-helical peptide.

KW Triple helix; collagen structure; homo-trimer; hetero-trimer.

OS Synthetic.

PN Key Location/Qualifiers

FT Modified-site 3 /label= 4Hyp

FT Modified-site 6 /label= 4Hyp

FT Modified-site 9
 FT /label= 4Hyp
 XX
 XX US5726243-A.
 XX
 PD 10-MAR-1998.
 XX
 XX 03-JUL-1996; 96US-0675140.
 XX
 XX 30-JUN-1993; 93US-0085633.
 PR 27-SEP-1995; 95US-0534342.
 PR 03-JUL-1996; 96US-0675140.
 XX
 XX (MIND) UNIV MINNESOTA.
 XX
 PI Fields GB;
 XX
 DR MPI; 1998-192815/17.
 XX
 PT Supported polypeptide(s) useful as intermediates - for
 PT triple-helical polypeptide(s) having collagen structures
 XX
 PS Disclosure; Column 3; 23pp; English.
 XX
 CC The invention relates to supported polypeptides of formula:
 CC A-B-B-(U)-(U)-L-R; where A = an N alpha-amino protecting group
 CC removable under non-acidic conditions; B = an amino acid having a single
 CC side-chain amino group protected with a group removable under non-strong
 CC acidic conditions; J = an amino acid that is capable of acting as a
 CC chromophore and is protected with a group capable of withstanding the
 CC conditions under which the protecting groups of A and B are removed; U =
 CC an amino acid; u = 0 or 1; j = 1 or more; L = a linker capable of being
 CC labilised using a non-strong acid mechanism, and R = a support material.
 CC The supported polypeptides are intermediates for triple-helical
 CC polypeptides. The triple-helical polypeptides are useful for studying the
 CC structure and biological activity of specific collagen sequences in homo-
 CC and hetero-trimeric forms. The present sequence represents a collagen
 CC sequence used in the synthesis of triple-helical peptides.
 CC
 SQ Sequence 24 AA;
 OY 1 GVGDKGNPGWPGAP 15
 DB 10 gvgdkgnpgwpgap 24
 SQ
 RESULT 10
 W13025
 ID W13025 standard; peptide: 27 AA.
 XX
 AC W13025;
 XX
 DT 30-APR-1997 (first entry)
 XX
 XX Collagen alpha(IV) chain containing triple helical polypeptide 4.
 XX
 XX Triple helical polypeptide; collagen; cell; adhesion; migration;
 KW promotion; alpha(IV) chain; solid phase synthesis; study;
 KW O-glycosidic bond; deuterium label; structure; biological activity;
 KW homotrimeric; heterotrimeric; nuclear magnetic resonance; NMR.
 XX
 OS Synthetic.
 XX
 FH Key Location/Qualifiers
 FT Modified-site 3
 FT /label= Hyp
 FT Modified-site 6
 FT /label= Hyp
 FT

FT Modified-site 9
 FT /label= Hyp
 FT Peptide 10..24
 FT /note= "collagen alpha(IV) chain residues
 FT 1263-1277"
 FT Modified-site 25
 FT /note= "linked via peptide bond to Pro24 of peptide
 FT comprising residues 1-24 of present
 FT sequence"
 FT Modified-site 26
 FT /note= "linked via peptide bond to Pro24 of peptide
 FT comprising residues 1-24 of present
 FT sequence"
 XX
 XX US576419-A.
 XX
 PD 19-NOV-1996.
 XX
 XX 30-JUN-1993; 93US-0085633.
 XX
 XX 30-JUN-1993; 93US-0085633.
 PR 27-SEP-1995; 95US-0534342.
 XX
 XX (MIND) UNIV MINNESOTA.
 XX
 PI Fields GB;
 XX
 DR MPI; 1997-011309/01.
 XX
 PT Solid phase synthesis of triple-helical branched polypeptide -
 PT under O-glycosidic bond and deuterium label retaining conditions,
 PT which may contain collagen cell adhesion sequences, useful for
 PT studying structure and biological activity of collagen
 XX
 PS Claim 52; Column 27; 25pp; English.
 XX
 CC The present sequence is that of triple helical polypeptide 4
 CC (THP-4), which includes a collagen cell adhesion sequence, known to
 CC promote adhesion and migration of various cells (residues 1263-1277
 CC of the collagen alpha(IV) chain), in its triple helix inducing
 CC region. THP-4 was prepared by solid phase synthesis under
 CC O-glycosidic bond and deuterium label retaining conditions, without
 CC strong acidolysis steps which affect O-glycosylation and isotopic
 CC labelling. THP can be used to study the structure and biological
 CC activity of particular collagens as homotrimeric or heterotrimeric
 CC forms, with the advantage that different adhesion sequences can be
 CC studied with a minimum of Gly-Pro-Hyp triplets, avoiding
 CC overlapping NMR resonances.
 CC
 SQ Sequence 27 AA;
 OY 1 GVGDKGNPGWPGAP 15
 DB 10 gvgdkgnpgwpgap 24
 SQ
 RESULT 11
 W46046
 ID W46046 standard; peptide: 39 AA.
 XX
 AC W46046;
 XX
 DT 23-DEC-1998 (first entry)
 XX
 XX Peptide intermediate used in synthesis of triple-helical peptide.
 DE Triple helix; collagen structure; homo-trimer; hetero-trimer.
 XX
 KW Triple helix; collagen structure; homo-trimer; hetero-trimer.
 XX

OS	Synthetic.
xx	
FH	Key
FT	Modified-site
FT	/Label= 4Hyp
FT	6
FT	Modified-site
FT	/Label= 4Hyp
FT	9
FT	Modified-site
FT	/Label= 4Hyp
FT	12
FT	Modified-site
FT	/Label= 4Hyp
FT	15
FT	Modified-site
FT	/Label= 4Hyp
FT	18
FT	Modified-site
FT	/Label= 4Hyp
FT	21
FT	Modified-site
FT	/Label= 4Hyp
FT	24
FT	Modified-site
FT	/Label= 4Hyp
FT	27
FT	Modified-site
FT	/note= "Lys(Boc)"
FT	29
FT	Modified-site
FT	/note= "Asp(OrBu)"
FT	30
FT	Modified-site
FT	/note= "Lys(Boc)"
FT	32
FT	Modified-site
FT	/note= "Asn(Trt)"
xx	
PN	US5726243-A.
PD	
XX	10-MAR-1998.
PF	
PP	03-JUL-1996;
PR	96US-0675140.
PR	30-JUN-1993;
PR	93US-0085633.
PR	27-SEP-1995;
PR	95US-0534342.
PA	03-JUL-1996;
PA	96US-0675140.
PI	(MINU) UNIV MINNESOTA.
PJ	
PI	Fields GB;
XX	
DR	WPI; 1998-192815/17.
PT	Supported polypeptide(s) useful as intermediates - for
PT	triple-helical polypeptide(s) having collagen structures
PS	
PS	Disclosure; Figure 3; 23pp; English.
xx	
CC	The invention relates to supported polypeptides of formula:
CC	A-B-B-(J)-(V)U-L-R; where A = an N alpha-amino protecting group
CC	removable under non-acidic conditions; B = an amino acid having a single
CC	side-chain amino group protected with a group removable under non-strong
CC	acidic conditions; J = an amino acid that is capable of acting as a
CC	chromophore and is protected with a group capable of withstanding the
CC	conditions under which the protecting groups of A and B are removed; U =
CC	an amino acid; u = 0 or 1; j = 1 or more; L = a linker capable of being
CC	labelled using a non-strong acid mechanism, and R = a support material.
CC	The supported polypeptides are intermediates for triple-helical
CC	polypeptides. The triple-helical polypeptides are useful for studying the
CC	structure and biological activity of specific collagen sequences in homo-
CC	and hetero-trimeric forms. The present sequence represents a peptide
CC	intermediate used in the synthesis of a triple-helical peptide.
SO	
Sequence	39 AA;
Query Match	100.0%; Score 92; DB 19; Length 39;
Best Local Similarity	100.0%; Pred. No. 7,le-07;
Matches	15; Conservative 0; Mismatches 0; Indels 0; Gaps 0
1 GYKGDKNRGPAGAP 15	

```

DB      25  gvkgdkgpnpwpgap 39

RESULT 12
ID      W46043
XX      W46043 standard; peptide; 39 AA.
XX      W46043;
AC
DT      23-DEC-1998 (first entry)
XX
XX      Collagen sequence synthesised as triple-helical peptide.
DE
XX      Triple helix; collagen structure; homo-trimer; hetero-trimer.
XX      Synthetic.
OS
XX      Synthetic.
FH      Key
FH      Modified-site
FT      3 Location/Qualifiers
FT      6 /label= 4HYP
FT      9 /label= 4HYP
FT      12 /label= 4HYP
FT      15 /label= 4HYP
FT      18 /label= 4HYP
FT      21 /label= 4HYP
FT      24 /label= 4HYP
FT      Modified-site
FT      Modified-site
FT      Modified-site
XX      24 /label= 4HYP
XX      US5726243-A.
XX      10-MAR-1998.
XX      03-JUL-1996; 96US-0675140.
XX      30-JUN-1993; 93US-0085633.
XX      27-SEP-1995; 95US-0534342.
XX      03-JUL-1996; 96US-0675140.
XX      (MINU ) UNIV MINNESOTA.
XX      Fields GB;
XX      WPI; 1998-192815/17.
XX      Supported polypeptide(s) useful as intermediates - for
XX      triple-helical polypeptide(s) having collagen structures
XX      Disclosure; Column 3; 23pp; English.
XX
XX      The invention relates to supported polypeptides of formula:
XX      A-B-B'(J)-U-L-R; where A = an N alpha-amino protecting group
XX      removable under non-acidic conditions; B = an amino acid having a single
XX      side-chain amino group protected with a group removable under non-strong
XX      acidic conditions; J = an amino acid that is capable of acting as a
XX      chromophore and is protected with a group capable of withstanding the
XX      conditions under which the protecting groups of A and B are removed; U =
XX      an amino acid; u = 0 or 1; j = 1 or more; L = a linker capable of being
XX      stabilised using a non-strong acid mechanism, and R = a support material.
XX      The supported polypeptides are intermediates for triple-helical
XX      polypeptides. The triple-helical polypeptides are useful for studying the
XX      structure and biological activity of specific collagen sequences in homo-
XX      and hetero-trimeric forms. The present sequence represents a collagen
XX      sequence used in the synthesis of triple-helical peptides.
XX
XX      Sequence 39 AA;
XX

```


CC antagonists may have neuroprotective; cytostatic; cardioactive;
CC immunomodulatory; muscular active general; vulnerary; gastrointestinal
CC activity; nephrotropic; antiinfective; gynecological; or antibacterial
CC activity. The invention also includes antibodies specific for the protein
CC or polynucleotide sequences. The lung cancer associated polynucleotide
CC sequences may be used for detection of lung cancer, chromosome
CC identification, as chromosome markers, and for numerous other diagnostic
CC or research purposes. The proteins may be used to treat disorders such as
CC neural, immune, muscular, reproductive, gastrointestinal, pulmonary,
CC cardiovascular, renal, and proliferative disorders. The proteins may also
CC be used in the treatment of wounds and infectious diseases.
CC Polynucleotide sequences F18425 - F18433 and peptide B58549 are used in
CC the course of the invention for the identification and characterisation
CC of the polynucleotide and protein sequences.

Sequence 406 AA:

Query Match 93.5%; Score 86; DB 21; Length 406;
Best Local Similarity 100.0%; Pred. No. 6.3e-05;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 VKGDKGNPQWPGAP 15
| | | | | | | | | | | | | | | | | |
Db 1 VKGDKGNPQWPGAP 14

RESULT 15

R23873
ID R23873 standard; Protein; 772 AA:

AC R23873:

DT 25-NOV-1992 (first entry)

XX Human alpha 5 (IV) of type IV collagen.

XX Mutations: Alport's syndrome; basement membranes; diabetes mellitus.

OS Homo sapiens.

XX Key Location/Qualifiers

FT Misc-difference 43..47 /note= "Interruption in Gly-X-Y sequence"

FT Misc-difference 159..160 /note= "Interruption in Gly-X-Y sequence"

FT Misc-difference 275..277 /note= "Interruption in Gly-X-Y sequence"

FT Misc-difference 334..336 /note= "Interruption in Gly-X-Y sequence"

FT Misc-difference 456..458 /note= "Interruption in Gly-X-Y sequence"

FT US5114840-A.

XX 19-MAY-1992.

XX 07-JUL-1989; 89US-0377238.

XX 07-JUL-1989; 89US-0377238.

XX (TRYG/) TRYGGVASON K.

XX Hostikka SL, Tryggvason K;

XX WPI: 1992-192174/23.

XX N-PSDB; Q24551.

PT Isolation of DNA encoding alpha-5(IV)polypeptide of type IV
PT collagen - to detect mutations in genes for alpha-5(IV) chain
PT which produce genetic or acquired basement membrane disorders
PT e.g. Alport's syndrome
XX

PS Disclosure; Fig 2; 14pp; English.

XX The sequence is that of the alpha 5(IV) polypeptide chain of human
XX type IV collagen, the major component of basement membranes. The
CC collagen contains the Gly-X-Y repeat coding sequence typical for
CC collagenous proteins at one end and a typical NC-domain coding
CC sequence at the other end. The sequence can be used to detect
CC mutations in individual genes specific for this chain which can,
CC directly or indirectly, produce several human diseases. It can
CC also be used to determine genetic, e.g. Alport's syndrome, or
CC acquired e.g. diabetes mellitus, disorders of the basement membrane,
CC and as probes or antibodies against these nucleotide sequences.
CC Gene fragments generated through amplifications from human genomic
CC or cloned DNA can also be used for detection and analysis of genes.

Sequence 772 AA:

Query Match 75.0%; Score 69; DB 13; Length 772;
Best Local Similarity 73.3%; Pred. No. 0.037;
Matches 11; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 1 GYKGDKNPQWPGAP 15
| : | : | | | | | | | | | | | | | |
Db 366 GYKGDKNPQWPGAP 380

RESULT 16

W09643
ID W09643 standard; Protein; 772 AA.

AC W09643:

DT 16-JUN-1997 (first entry)

XX Human type IV collagen alpha-5.

XX Collagen alpha5(IV); basement membrane; Alport's syndrome;
XX nephritis; kidney; renal failure; antibody; diagnosis; COL4A5 gene;
XX X chromosome.

OS Homo sapiens.

XX Key Location/Qualifiers

FT Key 1..543 /label= "Collagenous domain"

FT Domain /note= "collagenous domain contains Gly-X-Y
tripeptide repeats, interrupted at
positions 43-47, 159-160, 275-276, 334-335,
456-459"

FT Domain 544..772 /label= Non-collagenous domain

FT Peptide 742..751 /label= Immunogenic_peptide

FT /note= "peptide used to raise diagnostic
antibodies (Claim 1)"

FT US5593900-A.

XX 14-JAN-1997.

XX 07-JUL-1989; 89US-0377238.

XX 20-DEC-1990; 90US-0630563.

XX 07-JUL-1989; 89US-0377238.

XX 11-OCT-1994; 94US-0321084.

XX (HOST/) HOSTIKKA S L.

XX (HOYH/) HOYHTYA M.

XX (TRYG/) TRYGGVASON K.

XX Hostikka SL, Hoyhtya M, Tryggvason K;
XX

Query Match 100.0%; Score 92; DB 19; Length 39;
 Best Local Similarity 100.0%; Pred. No. 7.1e-07;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GVGKDGKNGPGWPGAP 15
 |||
 DB 25 GVGKDGKNGPGWPGAP 39

RESULT 13
 ID W13024 standard; peptide: 43 AA.
 XX W13024;
 AC
 XX
 DT 30-APR-1997 (first entry)
 XX
 DE Collagen alpha1(IV) chain containing triple helical polypeptide 2.
 XX
 XX Triple helical polypeptide; collagen; cell; adhesion; migration;
 KM promotion; alpha1(IV) chain; solid phase synthesis; study;
 KM O-glycosidic bond; deuterium label; structure; biological activity;
 KM homotrimeric; heterotrimeric; nuclear magnetic resonance; NMR.
 XX
 OS Synthetic.

Key Location/Qualifiers
 FT Modified-site 3 /label= Hyp
 FT Modified-site 6 /label= Hyp
 FT Modified-site 9 /label= Hyp
 FT Modified-site 12 /label= Hyp
 FT Modified-site 15 /label= Hyp
 FT Modified-site 18 /label= Hyp
 FT Modified-site 21 /label= Hyp
 FT Modified-site 24 /label= Hyp
 FT Modified-site 25..39 /label= Hyp
 FT Peptide /note= "collagen alpha1(IV) chain residues 1263-1277"
 FT Modified-site 40 /note= "linked via peptide bond to Pro39 of peptide comprising residues 1-39 of present sequence"
 FT Modified-site 41 /note= "linked via peptide bond to Pro39 of peptide comprising residues 1-39 of present sequence"

US5576419-A.
 PN 19-NOV-1996.
 PD 30-JUN-1993; 93US-0085633.
 PE 30-JUN-1993; 93US-0085633.
 PR 27-SEP-1995; 95US-0534342.
 XX (MINU) UNIV MINNESOTA.
 PA
 XX Fields GB;
 PI
 XX
 DR WPI; 1997-011309/01.
 XX
 PT Solid phase synthesis of triple-helical branched polypeptide -
 PT under O-glycosidic bond and deuterium label retaining conditions,

PT which may contain collagen cell adhesion sequences, useful for
 PT studying structure and biological activity of collagen
 XX
 XX
 PS Claim 48; Column 27; 25pp; English.

CC The present sequence is that of triple helical polypeptide 2
 CC (THP-2), which includes a collagen cell adhesion sequence, known to
 CC promote adhesion and migration of various cells (residues 1263-1277
 CC of the collagen alpha1(IV) chain), in its triple helix inducing
 CC region. THP-2 was prepared by solid phase synthesis under
 CC O-glycosidic bond and deuterium label retaining conditions, without
 CC strong acidolysis steps which affect O-glycosylation and isotopic
 CC labelling. THP can be used to study the structure and biological
 CC activity of particular collagens as homotrimeric or heterotrimeric
 CC forms, with the advantage that different adhesion sequences can be
 CC studied with a minimum of Gly-Pro-Hyp triplets, avoiding
 CC overlapping NMR resonances.

SQ Sequence 43 AA:

Query Match 100.0%; Score 92; DB 18; Length 43;
 Best Local Similarity 100.0%; Pred. No. 7.9e-07;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GVGKDGKNGPGWPGAP 15
 |||
 DB 25 GVGKDGKNGPGWPGAP 39

RESULT 14
 ID B58169 standard; Protein: 406 AA.
 XX B58169;
 AC
 XX
 DT 14-MAR-2001 (first entry)
 XX
 DE Lung cancer associated polypeptide sequence SEQ ID 507.
 XX
 XX Human; lung cancer associated protein; neuroprotective; cytostatic;
 KM cardioactive; immunomodulatory; muscular active; vulnary;
 KM gastrointestinal; nephrotoxic; antiinfective; gynecological;
 KM antibacterial; diagnosis; neural disorder; immune disorder; reproductive;
 KM proliferative disorder; wound healing; infectious disease.
 XX
 OS Homo sapiens.
 XX
 PN WO200055180-A2.
 XX
 PD 21-SEP-2000.
 XX
 PF 08-MAR-2000; 2000WO-US05918.
 XX
 PR 12-MAR-1999; 99US-0124270.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 PA (ROSE/) ROSEN C A.
 XX
 XX Ruben SM;
 PI
 XX
 DR WPI; 2000-587514/55.
 DR N-PSDB; F18045.
 XX
 PT Lung cancer associated gene sequences, referred to as lung cancer
 PT antigens, useful for treatment, prevention, and diagnosis of disorders
 PT such as lung cancer -
 XX
 XX
 PS Claim 11; Page 996-998; 1425pp; English.
 XX
 CC Polynucleotide sequences F17982 - F18424 encode human lung cancer
 CC associated proteins represented in B58106 - B58548. Lung cancer
 CC associated proteins and polynucleotide sequences, their agonists, and

CC antagonists may have neuroprotective; cytostatic; cardioactive;
 CC immunomodulatory; muscular active general; vulnary; gastrointestinal
 CC general; nephrotoxic; antiinfective; gynecological; or antibacterial
 CC activity. The invention also includes antibodies specific for the protein
 CC or polynucleotide sequences. The lung cancer associated polynucleotide
 CC sequences may be used for detection of lung cancer, chromosome
 CC identification, as chromosome markers, and for numerous other diagnostic
 CC or research purposes. The proteins may be used to treat disorders such as
 CC neural, immune, muscular, reproductive, gastrointestinal, pulmonary,
 CC cardiovascular, renal, and proliferative disorders. The proteins may also
 CC be used in the treatment of wounds and infectious diseases.
 CC polynucleotide sequences F18425 - F18433 and peptide B58549 are used in
 CC the course of the invention for the identification and characterisation
 CC of the polynucleotide and protein sequences.

CC Sequence 406 AA;

Query Match 93.5%; Score 86; DB 21; Length 406;
 Best Local Similarity 100.0%; Pred. No. 6.3e-05;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 VKGDKGNPGWPGAP 15
 1 VKGDKGNPGWPGAP 14

RESULT 15

R23873 R23873 standard; Protein: 772 AA.

AC R23873;

DT 25-NOV-1992 (first entry)

DE Human alpha 5 (IV) of type IV collagen.

XX Mutations; Alport's syndrome; basement membranes; diabetes mellitus.

OS Homo sapiens.

XX Key Location/Qualifiers

FT Misc-difference 43..47 /note= "interruption in Gly-X-Y sequence"

FT Misc-difference 159..160 /note= "interruption in Gly-X-Y sequence"

FT Misc-difference 275..277 /note= "interruption in Gly-X-Y sequence"

FT Misc-difference 334..336 /note= "interruption in Gly-X-Y sequence"

FT Misc-difference 456..458 /note= "interruption in Gly-X-Y sequence"

XX US5114840-A.

XX 19-MAY-1992.

XX 07-JUL-1989; 89US-0377238.

XX 07-JUL-1989; 89US-0377238.

XX (TRYG/) TRYGVASON K.

XX Hostikka SL, Trygvason K;

XX WPI: 1992-192174/23.

XX N-PSDB: Q24551.

PT Isolation of DNA encoding alpha-5(IV)polypeptide of type IV
 PT collagen - to detect mutations in genes for alpha-5(IV) chain
 PT which produce genetic or acquired basement membrane disorders
 PT e.g. Alport's syndrome
 XX

PS Disclosure; Fig 2; 14pp; English.

XX The sequence is that of the alpha 5(IV) polypeptide chain of human
 CC type IV collagen, the major component of basement membranes. The
 CC protein contains the Gly-X-Y repeat coding sequence typical for
 CC collagenous proteins at one end and a typical NC-domain coding
 CC sequence at the other end. The sequence can be used to detect
 CC mutations in individual genes specific for this chain which can,
 CC directly or indirectly, produce several human diseases. It can
 CC also be used to determine genetic, e.g. Alport's syndrome, or
 CC acquired e.g. diabetes mellitus, disorders of the basement membrane,
 CC and as probes or antibodies against these nucleotide sequences.
 CC Gene fragments generated through amplifications from human genomic
 CC or cloned DNA can also be used for detection and analysis of genes.

CC Sequence 772 AA;

Query Match 75.0%; Score 69; DB 13; Length 772;
 Best Local Similarity 73.3%; Pred. No. 0.037;
 Matches 11; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 1 GYKGDKNPGWPGAP 15
 1 GYKGDKNPGWPGAP 15

RESULT 16

W09643 W09643 standard; Protein: 772 AA.

AC W09643;

DT 16-JUN-1997 (first entry)

DE Human type IV collagen alpha-5.

XX Collagen alpha5(IV); basement membrane; Alport's syndrome;

XX nephritis; kidney; renal failure; antibody; diagnosis; COL4A5 gene;

OS X chromosome.

XX Homo sapiens.

XX Key Location/Qualifiers

FT Domain 1..543 /label= "collagenous domain"

FT Domain /note= "collagenous domain contains Gly-X-Y
 tripeptide repeats, interrupted at
 positions 43-47, 159-160, 275-276, 334-335,
 456-459"

FT Domain 544..772 /label= "Non-collagenous domain"

FT Peptide 742..751 /label= "Immunogenic peptide"

FT Peptide /note= "peptide used to raise diagnostic
 antibodies (Claim 1)"

XX US5593900-A.

XX 14-JAN-1997.

XX 07-JUL-1989; 89US-0377238.

XX 20-DEC-1990; 90US-0630563.

XX 07-JUL-1989; 89US-0377238.

XX 11-OCT-1994; 94US-0321084.

XX (HOST/) HOSTIKKA S L.

XX (HOYH/) HOYTIVA M.
 PA (TRYG/) TRYGVASON K.
 XX Hostikka SL, Hoyhtya M, Trygvason K;
 XX

PN US5731192-A.
 XX
 PD 24-MAR-1998.
 XX
 PF 23-JUN-1995; 95US-0494168.
 XX
 PR 27-AUG-1993; 93US-0112465.
 XX
 PA (UYVA) UNIV YALE.
 XX
 PI Reeders ST, Zhou J;
 XX
 DR WPI; 1998-216495/19.
 XX
 PT Nucleic acid encoding human alpha-6(IV) collagen - useful for, e.g.
 XX
 PS diagnosis or gene therapy of Alport's disease
 XX
 Example 1; Fig 3; 43pp; English.
 XX
 CC This sequence is the human alpha-1(IV) collagen protein which is used
 CC to analyse a novel human alpha-6(IV) collagen protein encoded by the
 CC COL4A6 gene. The alpha-6(IV) protein can be used for diagnosis or gene
 CC therapy of diseases associated with collagen type IV pathology,
 CC especially Alport's syndrome and associated diffuse leiomyomatosis. The
 CC polypeptide may also be used for generating monoclonal or polyclonal
 CC antibodies having specificity for the alpha-6(IV) polypeptide especially
 CC an antibody that is not crossreactive with other collagen proteins
 CC including alpha-1(IV), alpha-2(IV) and alpha-5(IV) collagens.
 CC
 SQ Sequence 532 AA;

Query Match 66.3%; Score 61; DB 19; Length 532;
 Best Local Similarity 66.7%; Pred. No. 0.36; Mismatches 2; Indels 0; Gaps 0;
 Matches 10; Conservative 3;

OY 1 GYVGDKGNPGWPGAP 15
 I | | | | | | | | | |
 DB 293 gdkgekspgfpgep 307

RESULT 19
 ID R71704 standard; protein; 1078 AA.
 XX
 AC R71704;
 XX
 DT 17-OCT-1995 (first entry)
 XX
 DE Collagen alpha 1 (III) chain precursor.
 XX
 KW Collagen; antibody; immunoassay; metabolism; diagnosis; monitoring;
 KW disorder; osteoporosis; metastatic progression; Paget's disease;
 KW hyperthyroidism; bone; resorption; rheumatoid arthritis;
 KW osteoarthritis; vasculitis syndrome.
 XX
 OS Homo sapiens.
 XX
 PN MO9508115-A.
 XX
 PD 23-MAR-1995.
 XX
 PF 19-SEP-1994; 94WO-DK00348.
 XX
 PR 17-SEP-1993; 93DK-0001040.
 XX
 PA (OSTE-) OSTEOMETER AS.
 XX
 PI Bonde M, Qvist P;
 XX
 DR WPI; 1995-131456/17.
 XX
 PT Assaying collagen fragments in body fluid by immunoassay - using

PT antibodies raised against synthetic peptide(s) contg. potential
 PT crosslinking sites, to diagnose and monitor disorders of collagen
 PT metabolism, e.g. osteoporosis.
 XX
 PS Disclosure (Appendix A); Page 55; 87pp; English.

XX
 CC Determination of collagen fragments in body fluids can be achieved
 CC by immunoassay using antibodies directed against synthetic peptides
 CC derived from collagen which contain sites of potential crosslinking.
 CC The method is used to diagnose and monitor treatment of disorders of
 CC collagen metabolism (degradation of type I collagen may indicate
 CC osteoporosis, metastatic progression, Paget's disease,
 CC hyperthyroidism or other conditions involving excessive bone
 CC resorption; degradation of type II collagen may indicate rheumatoid
 CC arthritis or osteoarthritis; and of type III collagen, vacuolitis
 CC syndrome). The method can also be used to assess the toxicity of a
 CC compound and to test drugs for their effect on collagen metabolism.
 XX
 SQ Sequence 1078 AA;

Query Match 65.2%; Score 60; DB 16; Length 1078;
 Best Local Similarity 78.6%; Pred. No. 1.1; Mismatches 3; Indels 0; Gaps 0;
 Matches 11; Conservative 0;

OY 1 GYVGDKGNPGWPGCA 14
 I | | | | | | | | | |
 DB 591 gpkgdgkspgpgga 604

RESULT 20
 ID Y96125 standard; Peptide; 1078 AA.
 XX

AC Y96125;
 XX
 DT 19-DEC-2000 (first entry)
 XX
 DE Collagen type III alpha-1.
 XX

KW Collagen type III; vasculitis syndrome; assay; diagnosis.

XX
 OS Homo sapiens.
 XX
 PN US6110689-A.
 XX
 PD 29-AUG-2000.
 XX
 PF 04-NOV-1997; 97US-0963825.
 XX
 PR 21-JAN-1994; 94US-0187319.
 XX
 PA (OSTE-) OSTEOMETER AS.
 XX
 PI Bonde M, Qvist P;
 XX
 DR WPI; 2000-586349/55.
 XX
 PT Assaying type I collagen fragments for diagnosing osteoporosis in
 PT postmenopausal woman, involves contacting body fluid with synthetic
 PT collagen peptide and antibody and quantifying by competitive binding
 PT assay -
 XX
 PS Disclosure; Column 46-51; 41pp; English.
 XX

CC The present sequence is that of human type III collagen alpha-1.
 CC The invention is based on the discovery of the presence of
 CC particular collagen fragments in body fluids of patients compared
 CC with those of healthy subjects. These fragments are generated
 CC upon collagen degradation and are partly characterised by the
 CC presence of potential sites for crosslinking. A method for
 CC assaying collagen fragments in a body fluid sample is based on the
 CC competitive binding to immunological binding partners of collagen

PR 30-NOV-1990; 900S-0621091.
 XX (UYVA) UNIV YALE.
 PA (UNIV) UNIV KANSAS MEDICAL CENT.
 XX
 XX Hudson BG, Reeders ST, Morrison KE;
 XX WPI; 1999-610317/52.
 DR N-PSDB; 228774.
 XX
 PT Isolated alpha 3 chain of type IV collagen polypeptide useful for
 PT diagnosis and treatment of Goodpasture syndrome -
 XX
 PS Claim 1; Column 31-34; 27pp; English.
 XX
 CC This sequence represents a recombinant bovine alpha3 chain of type IV
 CC collagen polypeptide. The sequence corresponds to the 238 amino acids
 CC of the C-terminal end of the triple helical domain and all 233 amino
 CC acids of the C-terminal non-collagenous domain. Alpha3 chain collagen
 CC polypeptides are useful for detecting Goodpasture antibodies in blood
 CC or tissue from a human patient and for treating Goodpasture syndrome,
 CC especially by neutralising the antibodies in the blood. The polypeptides
 CC also have a nephroretrophic activity.
 XX
 SQ Sequence 471 AA;

Query Match 64.1%; Score 59; DB 20; Length 471;
 Best Local Similarity 66.7%; Pred. No. 0.62;
 Matches 10; Conservative 2; Mismatches 3; Indels 0; Gaps 0;
 QY 1 GVGKDGKNGPWPAP 15
 ID 116 gmkgkkgngsfppp 130
 DB

RESULT 24
 Y56783
 ID Y56783 standard; Protein; 471 AA.
 XX
 AC Y56783;
 XX
 DT 27-MAR-2000 (first entry)
 XX
 DE Bovine alpha3 type IV collagen.
 XX
 KW Goodpasture syndrome; type IV collagen; alpha3 chain; bovine.
 XX
 OS Bos sp.
 XX
 PN US6007980-A.
 XX
 PD 28-DEC-1999.
 XX
 PF 07-OCT-1998; 98US-0167364.
 XX
 PR 30-NOV-1990; 90US-0621091.
 PR 07-MAR-1995; 95US-0399889.
 XX
 PA (UNIV) UNIV KANSAS MEDICAL CENT.
 PA (UYVA) UNIV YALE.
 XX
 PI Hudson BG, Reeders ST, Morrison KE;
 XX
 DR WPI; 2000-096371/08.
 DR N-PSDB; 246728.
 XX
 PT Diagnosing and treating Goodpasture syndrome using a peptide derived
 PT from type IV collagen -
 XX
 PS Disclosure: Columns 19-24; 26pp; English.
 XX
 CC The invention provides a method of detecting Goodpasture antibodies

CC in the fluid of a patient by contacting it with a peptide comprising at
 CC most 218 amino acids of the human alpha3 chain type IV collagen that
 CC contains the fragment shown in Y56785. The methods are useful for the
 CC diagnosis and treatment of Goodpasture syndrome. The present sequence
 CC represents the bovine alpha3 chain of type IV collagen.
 XX
 SQ Sequence 471 AA;

Query Match 64.1%; Score 59; DB 21; Length 471;
 Best Local Similarity 66.7%; Pred. No. 0.62;
 Matches 10; Conservative 2; Mismatches 3; Indels 0; Gaps 0;
 QY 1 GVGKDGKNGPWPAP 15
 ID 116 gmkgkkgngsfppp 130
 DB

RESULT 25
 W92297
 ID W92297 standard; peptide; 1288 AA.
 XX
 AC W92297;
 XX
 DT 28-APR-1999 (first entry)
 XX
 DE Mouse alpha-1 (XVIII) collagen chain common sequence MO18(common)28.
 XX
 KW Human; type XVIII collagen; liver disease; cirrhosis; detection;
 KW hepatocellular carcinoma; diagnosis.
 XX
 OS Mus sp.
 XX
 PN W09856399-A1.
 XX
 PD 17-DEC-1998.
 XX
 PF 12-JUN-1998; 98WO-US12327.
 XX
 PR 12-JUN-1997; 97US-0049369.
 XX
 PA (FIFT-) ACAD FINLAND.
 PA (FIBR-) FIBROGEN INC.
 PA (INRM) INST NAT SANTE & RECH MEDICALE.
 XX
 PI Clement B, Pihlajaniemi T, Rehn M;
 XX
 DR WPI; 1999-070292/06.
 XX
 PT Diagnosis and monitoring of liver disease by measuring collagen type
 PT XVIII levels - with elevated levels indicative of disease,
 PT especially cirrhosis or hepatocellular carcinoma
 XX
 PS Example 6; Fig 8; 56pp; English.
 XX
 CC A method has been developed for the detecting liver disease. The method
 CC comprises: (a) reacting a patient sample with antibodies (Ab) specific
 CC for collagen type XVIII (COL18); (b) measuring the amount of Ab-antigen
 CC complex (c) formed as indicator of the amount of COL18 present; (c)
 CC similar analysis of a non-diseased control; and (d) comparing the
 CC amounts of COL18 in the two samples to detect presence or progression of
 CC disease. Elevated levels of COL18 are: (i) indicative of disease,
 CC specifically cirrhosis; and (ii) predictive of the prognosis of disease,
 CC specifically hepatocellular carcinoma (there is a relationship between
 CC COL18 mRNA levels and tumour size and necrosis, and survival times are
 CC significantly higher in patients with higher COL18 levels). The method
 CC provides non-invasive, early and accurate diagnosis of liver disease.
 CC The present sequence represents the sequence common to mouse alpha-1
 CC (XVIII) collagen chain from the present invention.
 XX
 SQ Sequence 1288 AA;

Query Match 63.0%; Score 58; DB 20; Length 1288;
Best Local Similarity 66.7%; Pred. No. 2.5;
Matches 10; Conservative 1; Mismatches 4; Indels 0; Gaps 0;
QY 1 GYKGDKNPGMPGAP 15
| : | | | | | | | |
Db 300 gmkqkqepqpp 314

Search completed: May 23, 2001, 16:16:14
Job time: 85 sec

THIS PAGE BLANK (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 23, 2001, 16:15:10 ; Search time 11.65 seconds
(without alignments)
24.735 Million cell updates/sec

Title: US-09-529-691-1
Perfect score: 92
Sequence: 1 GVKGDKGNPGWGPAP 15

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 185757 seqs, 19210857 residues

Total number of hits satisfying chosen parameters: 185757

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/1aa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/1aa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/1aa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/1aa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/1aa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/1aa/backfillseq1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	92	100.0	15	3	US-08-394-748A-1
2	92	100.0	15	3	US-08-702-254A-1
3	92	100.0	15	5	PCT-US95-02478-1
4	92	100.0	24	1	US-08-534-342-3
5	92	100.0	24	1	US-08-675-140-3
6	92	100.0	39	1	US-08-534-342-2
7	92	100.0	39	1	US-08-675-140-2
8	66	71.7	39	1	US-08-534-342-5
9	66	71.7	39	1	US-08-534-342-6
10	66	71.7	39	1	US-08-675-140-5
11	66	71.7	39	1	US-08-675-140-6
12	61	65.2	532	1	US-08-494-168-9
13	60	65.2	1057	3	US-08-931-820-4
14	60	65.2	1078	4	US-08-963-825-21
15	59	64.1	471	2	US-08-399-889-24
16	59	64.1	471	2	US-09-167-364-24
17	56	60.9	489	2	US-08-794-795-7
18	56	60.9	489	4	US-09-249-200-7
19	56	60.9	518	1	US-08-392-367B-2
20	56	60.9	518	3	US-08-893-467A-2
21	56	60.9	546	1	US-08-494-168-10
22	55	59.8	171	4	US-09-011-735-2
23	55	59.8	351	4	US-09-011-735-1
24	55	59.8	1694	1	US-08-494-168-2
25	54	57.6	228	4	US-09-219-849-38
26	53	57.6	495	2	US-08-794-795-2
27	53	57.6	495	4	US-09-249-200-2

28	53	57.6	520	2	US-08-794-795-6	Sequence 6, Appl1
29	53	57.6	520	4	US-09-249-200-6	Sequence 6, Appl1
30	53	57.6	549	1	US-08-494-168-8	Sequence 8, Appl1
31	53	57.6	557	3	US-09-320-095-10	Sequence 10, Appl1
32	52	56.5	279	4	US-09-010-999-2	Sequence 2, Appl1
33	52	56.5	1057	3	US-08-931-820-1	Sequence 1, Appl1
34	52	56.5	1060	3	US-08-931-820-3	Sequence 3, Appl1
35	52	56.5	1418	4	US-08-963-825-20	Sequence 20, Appl1
36	52	56.5	1418	4	US-09-010-999-1	Sequence 1, Appl1
37	52	56.5	1442	2	US-08-316-650-12	Sequence 12, Appl1
38	52	56.5	1442	5	PCT-US95-02251-12	Sequence 12, Appl1
39	51	55.4	36	1	US-07-972-032-45	Sequence 45, Appl1
40	51	55.4	36	1	US-08-642-255-17	Sequence 17, Appl1
41	51	55.4	36	1	US-08-642-255-59	Sequence 59, Appl1
42	51	55.4	72	1	US-07-972-032-47	Sequence 47, Appl1
43	51	55.4	246	2	US-08-463-911-4	Sequence 4, Appl1
44	51	55.4	252	1	US-08-642-255-61	Sequence 61, Appl1
45	51	55.4	310	4	US-09-219-849-47	Sequence 47, Appl1

ALIGNMENTS

RESULT 1
US-08-394-748A-1
; Sequence 1, Application US/08394748A
; Patent No. 6013628
GENERAL INFORMATION:
APPLICANT: Skudlitz, Amy P.N.
APPLICANT: Furcht Leo T.
APPLICANT: Bailes, Mark
APPLICANT: Gregerson, Dale S.
APPLICANT: Agarwal, Anita
APPLICANT: Wright, Martha M.
APPLICANT: Mutrall, Shobana
TITLE OF INVENTION: Method for Treating Conditions of the Eye
TITLE OF INVENTION: Using Polypeptides
NUMBER OF SEQUENCES: 16
CORRESPONDENCE ADDRESS:
ADDRESSEE: Merchant & Gould
STREET: 3100 NO. 6013628west Center
CITY: Minneapolis
STATE: MN
COUNTRY: USA
ZIP: 55402
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/394,748A
FILING DATE: 27-FEB-1995
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/203,458
FILING DATE: 28-FEB-1994
ATTORNEY/AGENT INFORMATION:
NAME: Carter, Charles G.
REGISTRATION NUMBER: 35,093
REFERENCE/DOCKET NUMBER: 600.307US01
TELECOMMUNICATION INFORMATION:
TELEPHONE: 612-332-5300
TELEFAX: 612-332-9081
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 15 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
ORIGINAL SOURCE:
STRAIN: IV-H1
US-08-394-748A-1

Query Match 100.0%; Score 92; DB 3; Length 15;
Best Local Similarity 100.0%; Pred. No. 1.6e-07;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GVGKDGKNGPWPAP 15
Db 1 GVGKDGKNGPWPAP 15

RESULT 2
US-08-702-254A-1

Sequence 1, Application US/08702254A
Patent No. 6096863

GENERAL INFORMATION:

APPLICANT: FIELDS, GREGG B.

TITLE OF INVENTION: SELF-ASSEMBLING AMPHIPHILES FOR

TITLE OF INVENTION: CONSTRUCTION OF PEPTIDE SECONDARY STRUCTURES

NUMBER OF SEQUENCES: 3

CORRESPONDENCE ADDRESSES:

ADDRESSEE: MUETING, RAASCH & GERHARDT, P.A.

STREET: 119 NORTH FOURTH STREET

CITY: MINNEAPOLIS

STATE: MINNESOTA

COUNTRY: USA

ZIP: 55401

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/702,254A

FILING DATE: 23-AUG-1996

CLASSIFICATION: 530

ATTORNEY/AGENT INFORMATION:

NAME: MUETING, ANN M.

REGISTRATION NUMBER: 33,977

REFERENCE/DOCKET NUMBER: 110.00310101

TELECOMMUNICATION INFORMATION:

TELEPHONE: 612-305-1217

TELEFAX: 612-305-1228

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 15 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: peptide

US-08-702-254A-1

Query Match 100.0%; Score 92; DB 3; Length 15;
Best Local Similarity 100.0%; Pred. No. 1.6e-07;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GVGKDGKNGPWPAP 15
Db 1 GVGKDGKNGPWPAP 15

RESULT 3
PCT-US95-02478-1

Sequence 1, Application PC/TUS9502478

GENERAL INFORMATION:

APPLICANT:

TITLE OF INVENTION: Method for Treating Conditions

TITLE OF INVENTION: of the Eye Using Polypeptides

NUMBER OF SEQUENCES: 16

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/02478

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/203,458

FILING DATE: 28-FEB-1994

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 15 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: peptide

ORIGINAL SOURCE:

STRAIN: IV-HI

PCT-US95-02478-1

Query Match 100.0%; Score 92; DB 5; Length 15;
Best Local Similarity 100.0%; Pred. No. 1.6e-07;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GVGKDGKNGPWPAP 15
Db 1 GVGKDGKNGPWPAP 15

RESULT 4
US-08-534-342-3

Sequence 3, Application US/08534342

Patent No. 5576419

GENERAL INFORMATION:

APPLICANT: Fields, Gregg B.

TITLE OF INVENTION: Mild Solid-Phase Synthesis of Aligned,

NUMBER OF SEQUENCES: 13

CORRESPONDENCE ADDRESS:

ADDRESSEE: Merchant & Gould

STREET: 3100 No. 5576419west Center

CITY: Minneapolis

STATE: MN

COUNTRY: USA

ZIP: 55402

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/534,342

FILING DATE:

CLASSIFICATION: 530

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/085,633

FILING DATE: 30-JUN-1993

ATTORNEY/AGENT INFORMATION:

NAME: Mueting, Ann M.

REGISTRATION NUMBER: 33,977

REFERENCE/DOCKET NUMBER: 600.253-US-01

TELECOMMUNICATION INFORMATION:

TELEPHONE: 612-332-5300

TELEFAX: 612-332-9081

INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:

LENGTH: 24 amino acids

TYPE: amino acid

TOPOLOGY: unknown

US-08-534-342-3

Query Match 100.0%; Score 92; DB 1; Length 24;
Best Local Similarity 100.0%; Pred. No. 2.7e-07;

TYPE: amino acid
TOPOLOGY: unknown
US-08-675-140-2

Query Match 100.0%; Score 92; DB 1; Length 39;
Best Local Similarity 100.0%; Pred. No. 4.5e-07;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GYKGDKNPGWPGAP 15
DB 25 GYKGDKNPGWPGAP 39

RESULT 8
US-08-534-342-5
Sequence 5, Application US/08534342
Patent No. 5576419
GENERAL INFORMATION:
APPLICANT: Fields, Gregg B.
TITLE OF INVENTION: Mild Solid-Phase Synthesis of Aligned,
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:
ADDRESSEE: Merchant & Gould
STREET: 3100 No. 5576419west Center
CITY: Minneapolis
STATE: MN
COUNTRY: USA
ZIP: 55402

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/534,342
FILING DATE:
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/085,633
FILING DATE: 30-JUN-1993
ATTORNEY/AGENT INFORMATION:
NAME: Mueling, Ann M.
REGISTRATION NUMBER: 33,977
REFERENCE/DOCKET NUMBER: 600.253-US-01
TELECOMMUNICATION INFORMATION:
TELEPHONE: 612-332-5300
TELEFAX: 612-332-9081
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 39 amino acids
TYPE: amino acid
TOPOLOGY: unknown
US-08-534-342-5

Query Match 71.7%; Score 66; DB 1; Length 39;
Best Local Similarity 73.3%; Pred. No. 0.0024;
Matches 11; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 GYKGDKNPGWPGAP 15
DB 25 GYKGDKNPGWPGAP 39

RESULT 9
US-08-534-342-6
Sequence 6, Application US/08534342
Patent No. 5576419
GENERAL INFORMATION:
APPLICANT: Fields, Gregg B.
TITLE OF INVENTION: Mild Solid-Phase Synthesis of Aligned,
NUMBER OF SEQUENCES: 13

CORRESPONDENCE ADDRESS:
ADDRESSEE: Merchant & Gould
STREET: 3100 No. 5576419west Center
CITY: Minneapolis
STATE: MN
COUNTRY: USA
ZIP: 55402

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/534,342
FILING DATE:
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/085,633
FILING DATE: 30-JUN-1993
ATTORNEY/AGENT INFORMATION:
NAME: Mueling, Ann M.
REGISTRATION NUMBER: 33,977
REFERENCE/DOCKET NUMBER: 600.253-US-01
TELECOMMUNICATION INFORMATION:
TELEPHONE: 612-332-5300
TELEFAX: 612-332-9081
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 39 amino acids
TYPE: amino acid
TOPOLOGY: unknown
US-08-534-342-6

Query Match 71.7%; Score 66; DB 1; Length 39;
Best Local Similarity 73.3%; Pred. No. 0.0024;
Matches 11; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 GYKGDKNPGWPGAP 15
DB 25 GYKGDKNPGWPGAP 39

RESULT 10
US-08-675-140-5
Sequence 5, Application US/08675140
Patent No. 5726243
GENERAL INFORMATION:
APPLICANT: Fields, Gregg B.
TITLE OF INVENTION: Mild Solid-Phase Synthesis of Aligned,
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:
ADDRESSEE: Merchant & Gould
STREET: 3100 No. 5726243west Center
CITY: Minneapolis
STATE: MN
COUNTRY: USA
ZIP: 55402
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/675,140
FILING DATE: 03-JUL-1996
CLASSIFICATION: 525
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/085,633
FILING DATE: 30-JUN-1993
ATTORNEY/AGENT INFORMATION:
NAME: Mueling, Ann M.
REGISTRATION NUMBER: 33,977

REFERENCE/DOCKET NUMBER: 600,253-US-01
TELECOMMUNICATION INFORMATION:
TELEPHONE: 612-332-5300
TELEFAX: 612-332-9081
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 39 amino acids
TYPE: amino acid
TOPOLOGY: unknown
US-08-675-140-5

Query Match 71.7%; Score 66; DB 1; Length 39;
Best Local Similarity 73.3%; Pred. No. 0.0024;
Matches 11; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 GVGKDGKNGPMPGAP 15
||| |||||
Db 25 GVGKXGXGPGMPGAP 39

RESULT 11
US-08-675-140-6
Sequence 6, Application US/08675140
Patent No. 5726243
GENERAL INFORMATION:
APPLICANT: Fields, Gregg B.
TITLE OF INVENTION: Mild Solid-Phase Synthesis of Aligned,
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:
ADDRESSEE: Merchant & Gould
STREET: 3100 NO. 5726243west Center
CITY: Minneapolis
STATE: MN
COUNTRY: USA
ZIP: 55402
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/675,140
FILING DATE: 03-JUL-1996
CLASSIFICATION: 525
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/085,633
FILING DATE: 30-JUN-1993
ATTORNEY/AGENT INFORMATION:
NAME: Mueller, Ann M.
REGISTRATION NUMBER: 33,977
REFERENCE/DOCKET NUMBER: 600,253-US-01
TELECOMMUNICATION INFORMATION:
TELEPHONE: 612-332-5300
TELEFAX: 612-332-9081
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 39 amino acids
TYPE: amino acid
TOPOLOGY: unknown
US-08-675-140-6

Query Match 71.7%; Score 66; DB 1; Length 39;
Best Local Similarity 73.3%; Pred. No. 0.0024;
Matches 11; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 GVGKDGKNGPMPGAP 15
||| |||||
Db 25 GVGKXGXGPGMPGAP 39

RESULT 12

US-08-494-168-9
Sequence 9, Application US/08494168
Patent No. 5731192
GENERAL INFORMATION:
APPLICANT: Reeders, Stephen T.
TITLE OF INVENTION: Collagen COL4A6: Gene, Protein and Method
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 3000 K Street, N.W., Suite 500
CITY: Washington, D.C.
COUNTRY: USA
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/494,168
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/112,465
FILING DATE: 27-AUG-1993
ATTORNEY/AGENT INFORMATION:
NAME: Saxe, Bernhard D.
REGISTRATION NUMBER: 28,665
REFERENCE/DOCKET NUMBER: 40397/104/BABR
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202)672-5300
TELEFAX: (202)672-5399
TELEX: 904136
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 532 amino acids
TYPE: amino acid
TOPOLOGY: linear
US-08-494-168-9

Query Match 66.3%; Score 61; DB 1; Length 532;
Best Local Similarity 66.7%; Pred. No. 0.19;
Matches 10; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

OY 1 GVGKDGKNGPMPGAP 15
||| |||||
Db 293 GDKGKSGPGFPGEP 307

RESULT 13
US-08-931-820-4
Sequence 4, Application US/08931820
Patent No. 6010863
GENERAL INFORMATION:
APPLICANT:
TITLE OF INVENTION: Assay for collagen degradation
NUMBER OF SEQUENCES: 4
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/931,820
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: EP 96202596.1
FILING DATE:
INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:
LENGTH: 1057 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: Protein
HYPOTHETICAL: NO
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
TISSUE TYPE: Collagen type III
FEATURE:
NAME/KEY: Modified-site
LOCATION: 1055
OTHER INFORMATION: /label= Modified
OTHER INFORMATION: /note= "Ala may be Pro"
US-08-931-820-4

Query Match 65.2%; Score 60; DB 3; Length 1057;
Best Local Similarity 78.6%; Pred. No. 0.54;
Matches 11; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GYKGDKNPQWPGCA 14
1 11111 11111
Db 590 GPKDKGEPGPGCA 603

RESULT 14
US-08-963-825-21
Sequence 21, Application US/08963825
Patent No. 6110689
GENERAL INFORMATION:
APPLICANT: Qvist, Per
APPLICANT: Bonde, Martin
TITLE OF INVENTION: A Method for Assaying Collagen Fragments
TITLE OF INVENTION: In Body Fluids, A Test Kit and Means for Carrying Out the
TITLE OF INVENTION: Method and Use of the Method to Diagnose the Presence of
TITLE OF INVENTION: Disorders Associated with the Metabolism of
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESSEE: Darby & Darby PC
STREET: 805 Third Avenue
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10022

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/963,825
FILING DATE:
CLASSIFICATION: 436
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/187,319
FILING DATE: 21-JAN-1994
ATTORNEY/AGENT INFORMATION:
NAME: Gogoris, Adda C
REGISTRATION NUMBER: 29,714
REFERENCE/DOCKET NUMBER: 4305/08701
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-527-7700
TELEFAX: 212-753-6237
TELEX: 236687

INFORMATION FOR SEQ ID NO: 21:
SEQUENCE CHARACTERISTICS:
LENGTH: 1078 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
ORIGINAL SOURCE:

ORGANISM: Homo sapiens
IMMEDIATE SOURCE:
CLONE: COLLAGEN ALPHA 1 (III)
US-08-963-825-21

Query Match 65.2%; Score 60; DB 4; Length 1078;
Best Local Similarity 78.6%; Pred. No. 0.55;
Matches 11; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GYKGDKNPQWPGCA 14
1 11111 11111
Db 591 GPKDKGEPGPGCA 604

RESULT 15
US-08-399-889-24
Sequence 24, Application US/08399889B
Patent No. 5973120
GENERAL INFORMATION:
APPLICANT: Reeders, Stephen T
APPLICANT: Morrison, Karen E
APPLICANT: Hudson, Billy G
TITLE OF INVENTION: Alpha-3 Chain Type IV Collagen Polypeptides
FILE REFERENCE: 951263A
CURRENT APPLICATION NUMBER: US/08/399,889B
CURRENT FILING DATE: 1995-03-07
EARLIER APPLICATION NUMBER: 07/621091
EARLIER FILING DATE: 1990-11-30
NUMBER OF SEQ ID NOS: 25
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 24
LENGTH: 471.
TYPE: PRT
ORGANISM: Calf
US-08-399-889-24

Query Match 64.1%; Score 59; DB 2; Length 471;
Best Local Similarity 66.7%; Pred. No. 0.32;
Matches 10; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 GYKGDKNPQWPGAP 15
1:11111 11111
Db 116 GMYKKKNGSGFPGP 130

RESULT 16
US-09-167-364-24
Sequence 24, Application US/09167364
Patent No. 6007980
GENERAL INFORMATION:
APPLICANT: Reeders, Stephen T
APPLICANT: Morrison, Karen E
APPLICANT: Hudson, Billy G
TITLE OF INVENTION: Alpha-3 Chain Type IV Collagen Polypeptides
FILE REFERENCE: 951263B
CURRENT APPLICATION NUMBER: US/09/167,364
CURRENT FILING DATE: 1998-10-07
EARLIER APPLICATION NUMBER: 08/399889
EARLIER FILING DATE: 1995-03-07
NUMBER OF SEQ ID NOS: 25
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 24
LENGTH: 471
TYPE: PRT
ORGANISM: Calf
US-09-167-364-24

Query Match 64.1%; Score 59; DB 3; Length 471;
Best Local Similarity 66.7%; Pred. No. 0.32;
Matches 10; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 GVGKDKNGMPGAP 15
|:|:|:|:|:|:|
DB 116 GVGKDKNGSGPGRP 130

RESULT 17

US-08-794-795-7
Sequence 7, Application US/08794795
Patent No. 5916766

GENERAL INFORMATION:

APPLICANT: Elshourbagy, Nabli
APPLICANT: Adamou, John
APPLICANT: Gross, Mitchell
APPLICANT: Lyeko, Paul
TITLE OF INVENTION: Human Macro Scavenger Rec
TITLE OF INVENTION: eptor
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: Smithkline Beecham Corporation
STREET: 709 Swedeland Road
CITY: King of Prussia
STATE: PA
COUNTRY: USA
ZIP: 19406

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS

SOFTWARE: FASTSEQ for Windows Version 2.0

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/794,795

CLASSIFICATION: 435
FILING DATE: 04-FEB-1997

PRIOR APPLICATION DATA:
APPLICATION NUMBER: ATG50009P

FILING DATE: 22-MAY-1996

ATTORNEY/AGENT INFORMATION:
NAME: Han, William T

REGISTRATION NUMBER: 34,344
REFERENCE/DOCKET NUMBER: ATG50009

TELECOMMUNICATION INFORMATION:
TELEPHONE: 610-270-5219

TELEFAX: 610-270-4026

TELEX:
INFORMATION FOR SEQ ID NO: 7:

SEQUENCE CHARACTERISTICS:
LENGTH: 489 amino acids

TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein

US-08-794-795-7

Query Match 60.9%; Score 56; DB 2; Length 489;
Best Local Similarity 66.7%; Pred. No. 0.91;
Matches 10; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 GVGKDKNGMPGAP 15
|:|:|:|:|:|:|
DB 265 GVGKDKNGSGPGRP 279

RESULT 18

US-09-249-200-7
Sequence 7, Application US/09249200
Patent No. 6197931

GENERAL INFORMATION:

APPLICANT: ELSHOURBAGY, NABIL
APPLICANT: ADAMOU, JOHN
APPLICANT: GROSS, MITCHELL
APPLICANT: LYSKO, PAUL

TITLE OF INVENTION: HUMAN MARCO SCAVENGER RECEPTOR
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: Ratner & Prestia
STREET: P.O. Box 980
CITY: Valley Forge
STATE: PA
COUNTRY: USA
ZIP: 19482

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS

SOFTWARE: FASTSEQ for Windows Version 2.0

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/249,200

CLASSIFICATION: 2
FILING DATE: 12-FEB-1999

PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/794,795

FILING DATE: 04-FEB-1997
APPLICATION NUMBER: 60/017,699

FILING DATE: 23-MAY-1996

ATTORNEY/AGENT INFORMATION:
NAME: Prestia, Paul F

REGISTRATION NUMBER: 23,031
REFERENCE/DOCKET NUMBER: ATG-50009-1

TELECOMMUNICATION INFORMATION:
TELEPHONE: 610-407-0700

TELEFAX: 610-407-0700

TELEX: 846169

INFORMATION FOR SEQ ID NO: 7:

SEQUENCE CHARACTERISTICS:
LENGTH: 489 amino acids

TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein

US-09-249-200-7

Query Match 60.9%; Score 56; DB 4; Length 489;
Best Local Similarity 66.7%; Pred. No. 0.91;
Matches 10; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 GVGKDKNGMPGAP 15
|:|:|:|:|:|:|
DB 265 GVGKDKNGSGPGRP 279

RESULT 19

US-08-392-367B-2
Sequence 2, Application US/08392367B
Patent No. 5691197

GENERAL INFORMATION:

APPLICANT: Trygvsen, Karl
APPLICANT: Elomaa, Outi

APPLICANT: Kangas, Maarit
TITLE OF INVENTION: An isolated DNA Sequence for a

TITLE OF INVENTION: No. 5691197el Macrophage Receptor with

TITLE OF INVENTION: a Collagenous Domain and the

TITLE OF INVENTION: Polypeptide Chain Encoded by

NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fay, Sharpe, Beall, Fagan,

ADDRESSEE: Minnich & McKee
STREET: 1100 Superior Avenue

STREET: Suite 700
CITY: Cleveland

STATE: Ohio
COUNTRY: U.S.A.

ZIP: 44114-2518
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch,
MEDIUM TYPE: 720 Kb storable
COMPUTER: IBM PS/2, Model 35 SX
OPERATING SYSTEM: DOS 5.0
SOFTWARE: Word Perfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/392,367B
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Minnich, Richard J.
REGISTRATION NUMBER: 24,175
REFERENCE/DOCKET NUMBER: TRV 2 009
TELECOMMUNICATION INFORMATION:
TELEPHONE: (216) 861-5582
TELEFAX: (216) 241-1666
TELEX: (216) 980162
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 518 amino acids
TYPE: amino acid
STRANDEDNESS: Single
TOPOLOGY: Linear
US-08-392-367B-2

Query Match 60.9%; Score 56; DB 1; Length 518;
Best Local Similarity 66.7%; Pred. No. 0.96;
Matches 10; Conservative 1; Mismatches 4; Indels 0; Gaps 0;
QY 1 GVKGDKNPGWPGAP 15
||||:|||||
DB 294 GVKGDGKRGVGVGP 308

RESULT 20
US-08-893-467A-2
Sequence 2, Application US/08893467A
Patent No. 6063901
GENERAL INFORMATION:
APPLICANT: Tryggvason, Karl
APPLICANT: Eliomaa, Outi
APPLICANT: Kangas, Maarit
TITLE OF INVENTION: An Insolated DNA Sequence For a
Patent No. 6063901
TITLE OF INVENTION: No. 6063901el Macrophage Receptor with
TITLE OF INVENTION: a Collagenous Domain and the
TITLE OF INVENTION: Polypeptide Chain Encoded by
TITLE OF INVENTION: such a Sequence
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fay, Sharpe, Beall, Fagan,
ADDRESSEE: Minnich & McKee
STREET: 1100 Superior Avenue
STREET: Suite 700
CITY: Cleveland
STATE: Ohio
COUNTRY: U.S.A.
ZIP: 44114-2518
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch,
MEDIUM TYPE: 720 Kb storable
COMPUTER: IBM PS/2, Model 35 SX
OPERATING SYSTEM: DOS 5.0
SOFTWARE: Word Perfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/893,467A
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Minnich, Richard J.

REGISTRATION NUMBER: 24,175
REFERENCE/DOCKET NUMBER: TRV 2 009
TELECOMMUNICATION INFORMATION:
TELEPHONE: (216) 861-5582
TELEFAX: (216) 241-1666
TELEX: (216) 980162
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 518 amino acids
TYPE: amino acid
STRANDEDNESS: Single
TOPOLOGY: Linear
US-08-893-467A-2

Query Match 60.9%; Score 56; DB 3; Length 518;
Best Local Similarity 66.7%; Pred. No. 0.96;
Matches 10; Conservative 1; Mismatches 4; Indels 0; Gaps 0;
QY 1 GVKGDKNPGWPGAP 15
||||:|||||
DB 294 GVKGDGKRGVGVGP 308

RESULT 21
US-08-494-168-10
Sequence 10, Application US/08494168
Patent No. 5731192
GENERAL INFORMATION:
APPLICANT: Reeders, Stephen T.
APPLICANT: Zhou, Jing
TITLE OF INVENTION: Collagen COL4A6: Gene, protein and Method
TITLE OF INVENTION: of Detecting Collagen Deficiency
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 3000 K Street, N.W., Suite 500
CITY: Washington, D.C.
COUNTRY: USA
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/494,168
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/112,465
FILING DATE: 27-AUG-1993
ATTORNEY/AGENT INFORMATION:
NAME: Saxe, Bernhard D.
REGISTRATION NUMBER: 28,665
REFERENCE/DOCKET NUMBER: 40397/104/BAHR
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202)672-5300
TELEFAX: (202)672-5399
TELEX: 904136
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 546 amino acids
TYPE: amino acid
TOPOLOGY: Linear
US-08-494-168-10

Query Match 60.9%; Score 56; DB 1; Length 546;
Best Local Similarity 75.0%; Pred. No. 1;
Matches 9; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
QY 4 GVKGNPGWPGAP 15

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 23, 2001, 16:15:19 ; Search time 12.91 Seconds
(without alignments)
79,848 Million cell updates/sec

Title: US-09-529-691-1

Perfect score: 92

Sequence: 1 GVKGDKGNPMPGAP 15

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 198801 seqs, 68722935 residues

Total number of hits satisfying chosen parameters: 198801

Minimum DB seq length: 0
Maximum DB seq length: 200000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: PIR1:*
2: PIR2:*
3: PIR3:*
4: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	92	100.0	1669	1 CGH04B	collagen alpha 1(I)
2	78	84.8	1669	1 CGMS4B	collagen alpha 1(I)
3	72	78.3	356	2 S16907	collagen alpha 1(I)
4	69	75.0	1691	1 S22917	collagen alpha 5(I)
5	66	71.7	754	2 A35267	collagen alpha 5(I)
6	65	70.7	1702	2 A33526	collagen alpha 2(I)
7	64	69.6	1712	1 CGH02B	collagen alpha 2(I)
8	63	68.5	677	2 S23296	collagen alpha 2(I)
9	63	68.5	2844	2 A54849	collagen alpha 1(I)
10	62	67.4	290	2 A32249	collagen - sea urc
11	62	67.4	1414	1 S23809	collagen alpha 3(I)
12	62	67.4	1670	1 CGH03B	collagen alpha 3(I)
13	61	66.3	340	1 B46345	gene V protein - p
14	61	66.3	1603	2 S23810	collagen alpha 1(I)
15	61	66.3	1690	1 CGH01B	collagen alpha 4(I)
16	61	66.3	1763	2 S16366	collagen alpha 2(I)
17	60	65.2	310	2 S16096	collagen alpha 1(I)
18	60	65.2	623	2 A45137	collagen alpha 4(I)
19	60	65.2	636	2 S41067	collagen alpha 1(I)
20	60	65.2	886	2 S150694	collagen alpha 1(I)
21	60	65.2	1049	1 CGH07S	collagen alpha 1(I)
22	60	65.2	1466	1 CGH07L	collagen alpha 1(I)
23	60	65.2	1532	2 A61262	collagen alpha 1(I)
24	59.5	64.7	1752	2 T13990	collagen type IV a
25	59.5	64.7	1752	2 A45407	collagen alpha 3(I)
26	59	64.1	369	2 A33603	surfactant protein
27	59	64.1	471	2 A39024	collagen alpha 3(I)
28	59	64.1	1433	2 A46053	bullous pemphigoid
29	59	64.1	1744	2 S40991	collagen alpha 1(I)

30	58	63.0	285	2	S60598	collagen alpha cha
31	58	63.0	464	2	S59513	collagen II Al pro
32	58	63.0	1758	2	T29350	hypothetical prote
33	58	63.0	1759	2	T29351	collagen alpha 2(I)
34	57	62.0	298	2	JC1448	collagen col-34 -
35	57	62.0	650	2	T22002	hypothetical prote
36	57	62.0	920	2	A45748	collagen alpha 1(I)
37	57	62.0	1549	2	I48103	type VII collagen
38	56	60.9	299	2	T24833	hypothetical prote
39	56	60.9	380	2	A48295	collagen I - marin
40	56	60.9	518	2	A55840	macrophage bacteri
41	56	60.9	775	2	A61228	collagen alpha 2(I)
42	56	60.9	1142	2	JX0369	collagen alpha 1(I)
43	55.5	60.3	350	2	I73338	macrophage scaveng
44	55.5	60.3	458	2	B44407	macrophage scaveng
45	55	59.8	245	1	CIHUQC	complement subcomp

ALIGNMENTS

RESULT 1
CGH04B
collagen alpha 1(IV) chain precursor - human
N:Alternate names: procollagen alpha 1(IV) chain
C:Species: Homo sapiens (man)
C>Date: 28-May-1986 #sequence, revision 31-Dec-1992 #ext, change 07-Dec-1999
C:Accession: S16876; A32117; S02738; S00048; S25826; A23115; S00207; S39614; A02863;
R:Solinen, R.; Huotari, M.; Ganguly, A.; Prockop, D.J.; Trygvsen, K.
J. Biol. Chem. 264, 13565-13571, 1989
A:Title: Structural organization of the gene for the alpha-1 chain of human type IV c
A:Reference number: S16876; MUID:89340433
A:Accession: S16876
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-1669 <SO11>
A:Cross-references: EMBL:J04217; GB:J05039; NID:J180800; PIDN:AAA53097.1; PID:J180803
R:Poeschl, E.; Pollner, R.; Kuehn, K.
EMBO J. 7, 2687-2695, 1988
A:Title: The genes for the alpha1(IV) and alpha2(IV) chains of human basement membran
A:Reference number: S02738; MUID:89030632
A:Accession: S02738
A:Status: translation not shown
A:Molecule type: DNA
A:Residues: 1-6, 'L', '8-28 <POE>
A:Cross-references: EMBL:X12784; NID:J30072
R:Bräzel, D.; Oberbauer, I.; Dierlinger, H.; Babel, W.; Glanville, R.W.; Deutzmann, Eur. J. Biochem. 168, 529-536, 1987
A:Title: Completion of the amino acid sequence of the alpha1 chain of human basement
A:Reference number: S00048; MUID:88029471
A:Accession: S00048
A:Molecule type: mRNA
A:Residues: 1-318, 'A', 320-944 <BRA1>
A:Cross-references: EMBL:X05561; NID:J30066; PIDN:CAA29075.1; PID:J30067
A:Accession: S25826
A:Molecule type: protein
A:Residues: 271-318, 'A', 320-554 <BRA2>
R:Glanville, R.W.; Qian, R.Q.; Siebold, B.; Ristell, J.; Kuehn, K.
Eur. J. Biochem. 152, 213-219, 1985
A:Title: Amino acid sequence of the N-terminal aggregation and cross-linking region
A:Reference number: A23115; MUID:86004708
A:Accession: A23115
A:Molecule type: protein
A:Residues: 28-236, 'KE', 239-240, 'K', 242-243 <GLA>

A:Note: the amino end of the mature form is blocked
R:Solimieu, R.; Haka-Risku, T.; Prockop, D.J.; Tryggsason, K.
FEBS Lett. 225, 188-194, 1987
A:Title: Complete primary structure of the alpha(1)-chain of human basement membrane (type IV) collagen
A:Reference number: S00207; MUID:86083584
A:Accession: S00207
A:Molecule type: mRNA
A:Residues: 244-530 <S013>
A:Cross-references: EMBL:Y00706; NID:929548; PIDN:CA68698.1; PID:g29549
R:Edle, J.A.; Goldik, R.; Mann, K.; Kuehn, K.
EMBO J. 12, 4795-4802, 1993
A:Title: The alpha-1-beta-1 integrin recognition site of the basement membrane collagen
A:Reference number: S39614; MUID:94038963
A:Accession: S39614
A:Molecule type: protein
A:Residues: 371-554 <EBL>
R:Babel, W.; Gnanville, R.W.
Eur. J. Biochem. 143, 545-556, 1984
A:Title: Structure of human basement-membrane (type IV) collagen. Complete amino-acid sequence
A:Reference number: A02863; MUID:85003629
A:Accession: A02863
A:Molecule type: protein
A:Residues: 534-718, 'D', 720-836, 'Y', 838-841, 'P', 843-903, 'Q', 905-913, 'K', 915-997, 'K', 999-1000
A:Experimental source: placenta
R:Gnanville, R.W.; Rauter, A.
Hoppe-Seyler's Z. Physiol. Chem. 362, 943-951, 1981
A:Title: Peptid fragments of human placental basement-membrane collagens showing interrupted
A:Reference number: S16908; MUID:82005835
A:Accession: A58517
A:Molecule type: protein
A:Residues: 534-557, 'G', 539, 'G', 541-542, 'X', 544-553:1389-1405, 'XX', 1408-1409, 'X', 1411-1414
R:MacIntyre, R.S.; Benson, V.A.; Lovellio, K.T.; van der Rest, M.; Fietzek, P.P.
Biochemistry 22, 4940-4948, 1983
A:Title: Isolation and characterization of pepsin-solubilized human basement membrane (type IV) collagen
A:Reference number: S16910; MUID:84053346
A:Accession: S16910
A:Molecule type: protein
A:Residues: 534-537, 'G', 539, 'G', 541-542, 'G', 544-549, 939-940, 'W', 942-944, 'V', 946, 'X', 948-949
A:Experimental source: placenta
R:Philipianemi, T.; Tryggsason, K.; Myers, J.C.; Kurkinen, M.; Lebo, R.; Cheung, M.C.; F
J. Biol. Chem. 260, 7681-7687, 1985
A:Title: cDNA clones coding for the Pro-alpha-1(IV) chain of human type IV procollagen
A:Reference number: S01466; MUID:85207819
A:Accession: S01466
A:Molecule type: mRNA
A:Residues: 1256-1669 <PIH>
A:Cross-references: EMBL:M10940; NID:g180421; PIDN:AA52006.1; PID:g180424
R:Brinker, J.M.; Gude, L.U.; Loid, H.R.; Wang, S.Y.; Rosenbloom, J.; Kefalides, N.A.;
Proc. Natl. Acad. Sci. U.S.A. 82, 3649-3653, 1985
A:Title: Restricted homology between human alpha-1 type IV and other procollagen chains.
A:Reference number: S16879; MUID:85216555
A:Accession: S16879
A:Molecule type: mRNA
A:Residues: 1259-1669 <BRI>
A:Cross-references: EMBL:M11315; NID:g180817; PIDN:AA52042.1; PID:g180818
R:Oberbauer, I.; Laurent, M.; Schwartz, U.; Sakurai, Y.; Yamada, Y.; Vogeli, G.; Voss, R.
Eur. J. Biochem. 147, 217-224, 1985
A:Title: Amino acid sequence of the non-collagenous globular domain (NC1) of the alpha-1
A:Reference number: A02864; MUID:85127033
A:Accession: S19091
A:Molecule type: protein
A:Residues: 1435-1461, 'H', 1463-1482, 'X', 1484-1491, 1501-1514, 'X', 1516-1519, 1534-1553, 'X',
R:Stibbold, B.; Deitzmann, R.; Kuehn, K.
Eur. J. Biochem. 176, 617-624, 1988
A:Title: The arrangement of intra- and intermolecular disulfide bonds in the carboxyterm
A:Reference number: S02550; MUID:89005112
A:Contents: annotation; disulfide bonds
C:Genetics:
A:Gene: GDB:COL4A1
A:Cross-references: GDB:119791; OMIM:120130
A:Map position: 13q34-13q34
A:Introns: 28/3; 48/3; 78/3; 93/3; 108/3; 129/3; 147/3; 156/3; 184/3; 205/3; 217/3; 231/4

1: 731/3: 782/1, 820/1, 876/1, 906/1, 957/1, 990/1, 1020/1, 1066/3, 1109/1, 1136/1
C:Complex: type IV collagen is a heterotrimer of two alpha 1(IV) chains and one alpha
collagen among trimer amino-terminal domains (disulfide and desmosine cross-links),
r-trimer associations in the interrupted helical domain (with disulfide and desmosine
C:Function:
A:Description: structural component of extracellular basement membrane
C:Superfamily: collagen alpha 1(IV) chain
C:Keywords: basement membrane; blocked amino end; cell binding; coiled coil; duplicat
F:1-26/Domain: signal sequence #status predicted <SIG>
F:27-1669/Product: collagen alpha 1(IV) chain #status predicted <MAN>
F:29-162/Domain: amino-terminal nonhelical, 7S <7S>
F:163-1440/Domain: interrupted helical <COL>
F:414-452/Region: integrin binding #status experimental
F:591-599/Region: cell attachment (R-G-D) motif
F:917-919/Region: cell attachment (R-G-D) motif
F:968-970/Region: cell attachment (R-G-D) motif
F:1441-1669/Domain: carboxyl-terminal nonhelical, NC1 <NC1>
F:1451-1551/Domain: collagen IV carboxyl-terminal repeat <CTR>
F:1561-1665/Domain: collagen IV carboxyl-terminal repeat <CTR>
F:27/Modified site: blocked amino end (Aa) (In mature form) #status experimental
F:31,36,39,41,125,434,467,470/Disulfide bonds: interchain #status predicted
F:45,48,78,90,129,156,172,217,228,231,277,295,298,322,343,361,460,463,497,527,540,543,
1081,1084,1099,1117,1132,1150,1165,1182,1188,1206,1235,1265,1283,1304,1319,1328,
F:45,48,78,90,129,156,217,228,231,277,295,298,322,343,361,460,463,497,527,543,573,582
99,1117,1132,1150,1165,1182,1185,1188,1206,1235,1265,1283,1304,1319,1328,1340,1356,13
F:54,63,75,84,87,96,102,105,108,111,117,120,123,138,141,147,150,153,159,167,178,181,1
419,422,423,439,445,448,451,479,485,491,494,503,512,518,524,530,546,549,552,555,561,
9,745,748,751,754,763/Modified site: 4-hydroxyproline (Pro) #status experimental
F:126/Binding site: carbohydrate (Asn) (covalent) #status experimental
F:129/Modified site: allysine (Lys) #status predicted
F:122,540,947/Modified site: 5-hydroxylysine (Lys) #status atypical
F:272,645,839/Modified site: 4-hydroxyproline (Pro) #status atypical
F:446-447/Cleavage site: Gly-Ile (gelatinase B) #status predicted
F:766,775,784,787,790,796,799,804,810,816,822,834,860,863,869,872,875,887,890,893,899
23,1129,1138,1141,1159,1171,1176,1179,1194,1200,1203,1215,1224,1227,1244,1247,1250,12
41,1437/Modified site: 4-hydroxyproline (Pro) #status experimental
F:1120,1268/Modified site: 5-hydroxylysine (Lys) (partial) #status experimental
F:1120,1268/Binding site: carbohydrate (Lys) (covalent) (partial) #status experimental
F:1124,1424/Modified site: 3-hydroxyproline (Pro) #status absent
F:1392,1395,1398,1404/Modified site: 4-hydroxyproline (Pro) #status experimental
F:1460-1548,1493-1551/Disulfide bonds: (or 1460-1551, 1493-1548) #status predicted
F:1505-1511,1616-1622/Disulfide bonds: #status predicted
F:1570-1662,1604-1665/Disulfide bonds: (or 1570-1665, 1604-1662) #status predicted

Query Match 100.0%; Score 92; DB 1; Length 1669;
Best Local Similarity 100.0%; Pred. No. 1,2e-05;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GVKGDKGNPWGAP 15
DB 1263 GVKGDKGNPWGAP 1277

RESULT 2
CGMS4B
collagen alpha 1(IV) chain precursor - mouse
C:Species: Mus musculus (house mouse)
C:Date: 28-May-1966 #sequence.revision 31-Dec-1992 #text.change 16-Jun-2000
C:Accession: A33525; S0154; A28066; A02664; A26363; A29301; S19079; A32003; A31766;
R:Murthukumaran, G.; Blumberg, B.; Kurkinen, M.
J. Biol. Chem. 264, 6310-6317, 1989
A:Title: The complete primary structure for the alpha-1-chain of mouse collagen IV. D
A:Accession: A33525; MUID:89197932
A:Accession: A33525
A:Molecule type: mRNA
A:Residues: 1-1669 <MUT>
A:Cross-references: EMBL:J04694; NID:g556296; PIDN:AAA50292.1; PID:g556297
R:Wood, L.; Theriault, N.; Vogel, G.
FEBS Lett. 227, 5-8, 1988
A:Title: cDNA clones completing the nucleotide and derived amino acid sequence of the
A:Reference number: S01454; MUID:88112221
A:Accession: S01454

A:Molecule type: mRNA
 A:Residues: 1-185, 'L', 187-318, 'S', 320-368, 'L', 370-402, 'F', 404-480, 'L', 482-492, 'H', 494-71
 A:Cross-references: EMBL:X06777
 R:Kilien, P.D.; Burdello, P.; Sakurai, Y.; Yamada, Y.
 J. Biol. Chem. 263, 8706-8709, 1988
 A:Title: Structure of the amino-terminal portion of the murine alpha-1(IV) collagen chain
 A:Reference number: A28066; MUID:88243724
 A:Accession: A28066
 A:Molecule type: protein
 A:Residues: 1-129 <K11>
 A:Cross-references: EMBL:D03758; NID:g192669; PIDN:AAA37439.1; PID:g192670
 R:Oberdaemmer, I.; Laurent, M.; Schwartz, U.; Sakurai, Y.; Yamada, Y.; Vogell, G.; Voss, E.
 Eur. J. Biochem. 147, 217-224, 1985
 A:Title: Amino acid sequence of the non-collagenous globular domain (NC1) of the alpha-1
 A:Reference number: A02864; MUID:85127033
 A:Accession: A02864
 A:Molecule type: mRNA
 A:Residues: 1276-1669 <OBE>
 A:Cross-references: EMBL:X02201; NID:g50233; PIDN:CAA26132.1; PID:g1333876
 R:Nath, P.; Laurent, M.; Horn, E.; Sobel, M.E.; Zon, G.; Vogell, G.
 Gene 43, 301-304, 1986
 A:Title: Isolation of an alpha-1 type-IV collagen cDNA clone using a synthetic oligodeox
 A:Reference number: A25636; MUID:86301886
 A:Accession: A25636
 A:Molecule type: mRNA
 A:Residues: 1149-1396, 'S', 1398-1424 <NAT>
 A:Cross-references: EMBL:M14042; NID:g192286; PIDN:AAA37342.1; PID:g192287
 A:Note: the authors translated the codon CAG for residue 1374 as Arg
 R:Kurlinen, M.; Condon, M.R.; Blumberg, B.; Barlow, D.P.; Quiñones, S.; Saus, J.; Pihla
 J. Biol. Chem. 262, 8496-8499, 1987
 A:Title: Extensive homology between the carboxyl-terminal peptides of mouse alpha-1(IV)
 A:Reference number: A94680; MUID:87250460
 A:Accession: A29301
 A:Molecule type: mRNA
 A:Residues: 1441-1669 <KUR>
 A:Cross-references: EMBL:M15832; NID:g192282; PIDN:AAA37340.1; PID:g387115
 R:Kilien, P.D.; Burdello, P.D.; Martin, G.R.; Yamada, Y.
 J. Biol. Chem. 263, 12310-12314, 1988
 A:Title: Characterization of the promoter for the alpha-1(IV) collagen gene. DNA sequenc
 A:Reference number: S19079; MUID:88315019
 A:Accession: S19079
 A:Molecule type: DNA
 A:Residues: 1-28 <K12>
 A:Cross-references: EMBL:J03944; NID:g192673; PIDN:AAA37442.1; PID:g466503
 R:Kayes, P.; Wood, L.; Thieriault, N.; Kurlinen, M.; Vogell, G.
 J. Biol. Chem. 263, 19274-19277, 1988
 A:Title: Head-to-head arrangement of murine type IV collagen genes.
 A:Reference number: A92702; MUID:89066738
 A:Accession: A32003
 A:Molecule type: DNA
 A:Residues: 1-28 <KAV>
 A:Cross-references: EMBL:J04448; NID:g192666; PIDN:AAA37437.1; PID:g450449
 R:Burdello, P.D.; Martin, G.R.; Yamada, Y.
 Proc. Natl. Acad. Sci. U.S.A. 85, 9679-9682, 1988
 A:Title: Alpha1(IV) and alpha2(IV) collagen genes are regulated by a bidirectional prom
 A:Reference number: A94220; MUID:89071159
 A:Accession: A31766
 A:Molecule type: DNA
 A:Residues: 1-28 <BUR>
 A:Cross-references: EMBL:M23333; NID:g340878; PIDN:AAA51625.1; PID:g555668
 R:Sakurai, Y.; Sullivan, M.; Yamada, Y.
 J. Biol. Chem. 261, 6654-6657, 1986
 A:Title: Alpha-1 type IV collagen gene evolved differently from fibrillar collagen genes
 A:Reference number: S19094; MUID:86196099
 A:Accession: S19094
 A:Molecule type: DNA
 A:Residues: 1110-1135, 1189-1316, 1342-1383, 1418-1487 <SAK>
 A:Cross-references: EMBL:M13027
 R:Schuppan, D.; Timpl, R.; Glanville, R.W.
 FEBS Lett. 115, 297-300, 1980
 A:Title: Discontinuities in the triple helical sequence Gly-X-Y of basement membrane (ty
 A:Reference number: S16909; MUID:80246483
 A:Accession: S16909

A:Molecule type: protein
 A:Residues: 940-946, 'G', 948-949, 'G', 951-955, 'G', 957-1213-1228, 'X', 1230-1234, 'P', 1236-
 R:Schuppan, D.; Glanville, R.W.; Timpl, R.
 Eur. J. Biochem. 123, 505-512, 1982
 A:Title: Covalent structure of mouse type-IV collagen. Isolation, order and partial a
 A:Reference number: A25991; MUID:82186723
 A:Accession: A25991
 A:Molecule type: protein
 A:Residues: 940-946, 'X', 948-949, 'X', 951-955, 'X', 957-964, 'X', 966-991, 'X', 993-1003, 'X',
 61, 'X', 1063-1065, 'X', 1067-1080, 'X', 1082-1083, 'X', 1085-1106, 'X', 1108-1115, 'DE', 1118-11
 A:Accession: B25991
 A:Molecule type: protein
 A:Residues: 1173-1181, 'X', 1183-1184, 'X', 1186-1187, 'X', 1189-1205, 'Q', 1207, 'XE', 1210-12
 3, 'SP', 1266, 'IT', 1269, 'SK', 1272, 'DM', 1275, 'L', 1277-1282, 1316-1318, 'X', 1320-1327, 'X', 1
 R:Reber, S.; Engel, J.; Wiedemann, H.; Glanville, R.W.; Timpl, R.
 Eur. J. Biochem. 139, 401-410, 1984
 A:Title: Subunit structure and assembly of the globular domain of basement-membrane c
 A:Reference number: S17801; MUID:84132058
 A:Accession: S17801
 A:Molecule type: protein
 A:Residues: 1435-1443 <WEB>
 C:Genetics: 28/3; 48/3; 78/3; 93/3; 108/3; 129/3
 A:Note: the list of introns may be incomplete
 C:Superfamily: collagen alpha 1(IV) chain
 C:Keywords: basement membrane; cell binding; coiled coll.; duplication; extracellular
 F:1-27/Domain: signal sequence #status predicted <Sig>
 F:28-1669/Product: collagen alpha 1(IV) chain #status predicted <Mat>
 F:28-162/Domain: 75 <7SD>
 F:163-1440/Domain: collagenous, triple helix <COL>
 F:597-599/Region: cell attachment (R-G-D) motif
 F:781-783/Region: cell attachment (R-G-D) motif
 F:917-919/Region: cell attachment (R-G-D) motif
 F:968-970/Region: cell attachment (R-G-D) motif
 F:1441-1669/Domain: carboxyl-terminal nonhelical, NC1 <NC1>
 F:1441-1552/Region: duplication
 F:1553-1669/Region: duplication
 F:31,36,39,41,43,467,470/Disulfide bonds: interchain #status predicted
 F:126/Binding site: carboxylate (Asn) (covalent) #status predicted
 F:971,974,977,986,989,1001,1007,1019,1022,1031,1037,1040,1055,1060,1063,1075,1078,109
 92,1298,1310,1313,1322,1337,1349,1349,1422,1425,1431,1437,1440/Modified site: hydroxy
 F:1214,1424/Modified site: 4-hydroxyproline (Pro) #status experimental
 F:1304/Modified site: 5-hydroxylysine (Lys) #status experimental
 F:1505-1511,1616-1622/Disulfide bonds: #status predicted

Query Match 84.8%; Score 78; DB 1; Length 1669;
 Best Local Similarity 86.7%; Pred. No. 0.0014;
 Matches 13; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GVGKDGKNGMPGAP 15
 DB 1263 GPRGDKNGQMGPCAP 1277

RESULT 3
 S16907
 collagen alpha 1(IV) chain - bovine (fragments)
 C:Species: Bos primigenius taurus (cattle)
 C:Date: 04-Dec-1992 #sequence, revision 05-Apr-1995 #text-change 19-Oct-1995
 C:Accession: A39474; S16907; S18432; A39419
 R:Gunwar, S.; Noelken, M.E.; Hudson, B.G.
 J. Biol. Chem. 266, 14088-14094, 1991
 A:Title: Properties of the collagenous domain of the alpha3(IV) chain, the coopastur
 le helical structure and noncollagenous domain.
 A:Reference number: A39474; MUID:91310700
 A:Accession: A39474
 A:Molecule type: protein
 A:Residues: 1-9 <GV2>
 R:Schuppan, D.; Glanville, R.W.; Timpl, R.; Dixlt, S.N.; Kang, A.H.
 Biochem. J. 220, 227-233, 1984
 A:Title: Sequence comparison of pepsin-resistant segments of basement-membrane collag
 A:Reference number: S16907; MUID:84256630

A:Accession: S16907
 A:Molecule type: protein
 A:Note: the authors translated the codon GCC for residue 115 as Val
 A:Residues: 4-8, 'E', 10-68; 69-158; 159-321; 322-349 <SC2>
 R:Butkowski, R.J.; Langeveld, J.P.M.; Mieslander, J.; Hamilton, J.; Hudson, B.G.
 J. Biol. Chem. 262, 7874-7877, 1987
 A:Title: Localization of the Goodpasture epitope to a novel chain of basement membrane
 A:Reference number: S18432; MUID:87222419
 A:Accession: S18432
 A:Molecule type: protein
 A:Residues: 337-347, 'FL', 350-353 <BU2>
 R:Gunnar, S.; Ballestier, F.; Kalluri, R.; Timoneda, J.; Chonko, A.M.; Edwards, S.T.; Noe
 J. Biol. Chem. 266, 15318-15324, 1991
 A:Title: Glomerular basement membrane. Identification of dimeric subunits of the noncoll
 A:Reference number: A39419; MUID:91332055
 A:Accession: A39419
 A:Molecule type: protein
 A:Residues: 337-347, 'FL', 350-356 <GU3>
 C:Superfamily: collagen alpha 1(IV) chain
 C:Keywords: basement membrane; cell binding; coiled coil; disulfide bond; extracellular
 F:31,34,37,46,61,69,78,84,87,102,110,122,125,137,140,143,149,155,158,161,164,185,188,199
 proline (Pro) #status experimental

Query Match 78.3%; Score 72; DB 2; Length 356;
 Best Local Similarity 66.7%; Pred. No. 0.0033;
 Matches 10; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

OY 1 GVKGDKNPCWPCAP 15
 ||:|||||||
 Db 248 GLAGERGNPCWPCPT 262

RESULT 4

S22917
 collagen alpha 5(IV) chain precursor, renal splice form - human
 N:Alternative names: procollagen alpha 5(IV) chain
 N:Contains: collagen alpha 5(IV) chain precursor, leukocyte splice form
 C:Species: Homo sapiens (man)
 C:Date: 30-Sep-1993 #sequence revision 27-Feb-1997 #text change 21-Jul-2000
 C:Accession: S22917; A54365; A57079; A37122; I54317; A34850; S18850; I56971; I76598; A35
 R:Zhou, J.; Hertz, J.M.; Leinonen, A.; Trygvason, K.
 J. Biol. Chem. 267, 12475-12481, 1992
 A:Title: Complete amino acid sequence of the human alpha-5(IV) collagen chain and identifi
 A:Note: syndrome patient.
 A:Reference number: S22917; MUID:92316923
 A:Accession: S22917
 A:Molecule type: mRNA
 A:Residues: 1-967 <ZHO>
 A:Cross-references: GB:M90464; NID:9180826; PIDN:AAA52046.1; PID:9553234
 R:Zhou, J.; Leinonen, A.; Trygvason, K.
 J. Biol. Chem. 269, 6608-6614, 1994
 A:Title: Structure of the human type IV collagen COL4A5 gene.
 A:Reference number: A54365; MUID:94165049
 A:Accession: A54365
 A:Molecule type: DNA
 A:Molecule type: DNA
 A:Residues: 1-922 <ZHO>
 A:Cross-references: GB:U04470; NID:9463378; GB:U04520; NID:9463428; PIDN:ACC27816.1; PID
 R:Zhou, J.; Mochizuki, T.; Smeets, H.; Antignac, C.; Laurila, P.; de Paape, A.; Trygvas
 Science 261, 1167-1169, 1993
 A:Title: Deletion of the paired alpha5(IV) and alpha6(IV) collagen genes in inherited sm
 A:Reference number: A57079; MUID:93361972
 A:Accession: A57079
 A:Molecule type: DNA
 A:Residues: 1-27 <ZHO>
 A:Cross-references: GB:Z37153; NID:9587203; PIDN:CAA85512.1; PID:9587204
 R:Philafanteml, T.; Pohjalainen, E.R.; Myers, J.C.
 J. Biol. Chem. 265, 13758-13766, 1990
 A:Title: Complete primary structure of the triple-helical region and the carboxyl-termin
 A:Reference number: A37122; MUID:90337990
 A:Accession: A37122
 A:Molecule type: mRNA
 A:Residues: 84-439, 'GS', 442-624, 'LAIO', 629-666, 'FR', 669-887, 'R', 889-1264, 1271-1691 <PIH>
 A:Cross-references: GB:J05558; EMBL:M88526; NID:91314209

A:Note: submitted to the EMBL Data Library, February 1991
 A:Note: the authors translated the codon GCC for residue 115 as Val
 R:Renteri, A.; Seri, M.; Myers, J.C.; Philafanteml, T.; Massella, L.; Rizzoni, G.; De
 Hum Mol. Genet. 1, 127-129, 1992
 A:Title: De novo mutation in the COL4A5 gene converting glycine 325 to glutamic acid
 A:Reference number: I54317; MUID:93244772
 A:Accession: I54317
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 313-324, 'E', 326-330 <REN>
 A:Cross-references: GB:S59334; NID:9299946; PIDN:RAD13909.1; PID:94261609
 R:Hostikka, S.L.; Eddy, R.L.; Byers, M.G.; Hoeyhtyae, M.; Shows, T.B.; Trygvason, K.
 Proc. Natl. Acad. Sci. U.S.A. 87, 1606-1610, 1990
 A:Title: Identification of a distinct type IV collagen alpha chain with restricted ki
 A:Reference number: A34850; MUID:90160375
 A:Accession: A34850
 A:Molecule type: mRNA
 A:Residues: 914-1264, 1271-1691 <ROS>
 A:Cross-references: EMBL:M31115; NID:9180824; PIDN:AAA52045.1; PID:9180825
 R:Zhou, J.; Hostikka, S.L.; Chow, L.T.; Trygvason, K.
 Genomics 9, 1-9, 1991
 A:Title: Characterization of the 3' half of the human type IV collagen alpha-5 gene t
 A:Reference number: A37969; MUID:91169491
 A:Accession: A37969
 A:Molecule type: DNA
 A:Residues: 924-1264, 1271-1691 <ZH3>
 A:Cross-references: EMBL:M63455; EMBL:M63456; EMBL:M63457; EMBL:M63458; EMBL:M63459;
 B: EMBL:M63470; EMBL:M63471; EMBL:M63472; EMBL:M63473; NID:9177922; PIDN:AAA51558.1;
 R:Guo, C.; Van Damme, B.; Van Damme-Lombaerts, R.; Van den Berghe, H.; Cassiman, J.J.
 Kidney Int. 44, 1316-1321, 1993
 A:Title: Differential splicing of COL4A5 mRNA in kidney and white blood cells: a comp
 A:Reference number: I56971; MUID:94133540
 A:Accession: I56971
 A:Status: translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1258-1276 <GUO1>
 A:Cross-references: GB:S69168; NID:9545095; PIDN:AC60612.1; PID:9545096
 A:Note: kidney splice form
 A:Accession: I76598
 A:Status: translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1284-1291, 'TETGYLACLV', <GUO2>
 A:Cross-references: GB:S69169; NID:9545097; PIDN:AC60613.1; PID:9545098
 A:Note: frameshift mutation in patient with Alport syndrome
 R:Myers, J.C.; Jones, T.A.; Pohjalainen, E.R.; Kadri, A.S.; Goddard, A.D.; Sheer, D.;
 Am. J. Hum. Genet. 46, 1024-1033, 1990
 A:Title: Molecular cloning of alpha5(IV) collagen and assignment of the gene to the r
 A:Reference number: A35335; MUID:90252791
 A:Accession: A35335
 A:Status: nucleic acid sequence not shown
 A:Molecule type: mRNA
 A:Residues: 1448-1477 <MYE>
 R:Nakazato, H.; Hattori, S.; Ushijima, T.; Matsuura, T.; Kotlabashi, Y.; Takada, T.;
 Kidney Int. 46, 1307-1314, 1994
 A:Title: Mutations in the COL4A5 gene in Alport syndrome: a possible mutation in prlm
 A:Reference number: I56975; MUID:95156893
 A:Accession: I56975
 A:Status: translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1395-1602 <NAK>
 A:Cross-references: GB:S75903; NID:9913882; PIDN:AAB3374.1; PID:9913883
 A:Note: premature termination mutation from a patient with Alport syndrome; one other
 R:Lemink, H.H.; Schroeder, C.H.; Brunner, H.G.; Neelen, M.R.; Zhou, J.; Trygvason, K.
 Genomics 17, 485-489, 1993
 A:Title: Identification of four novel mutations in the COL4A5 gene of patients with A
 A:Reference number: I54188; MUID:94010948
 A:Accession: I54188
 A:Status: translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1604-1607, 'VHDAYKC', <LEK>
 A:Cross-references: GB:S65767; NID:9425563; PIDN:AAD13967.1; PID:94261667
 A:Note: frameshift mutation from a patient with Alport syndrome; five other mutations
 C:Comment: Prolines and lysines at the third position of the tripeptide repeating unit

Proc. Natl. Acad. Sci. U.S.A. 85, 9679-9682, 1988
 A:Title: Alpha1(IV) and alpha2(IV) collagen genes are regulated by a bidirectional promoter
 A:Reference number: A94220; MUID:89071759
 A:Accession: B31766
 A:Molecule type: DNA
 A:Residues: 1-60

 A:Cross-references: EMBL:M23333
 R:Weber, S.; Engel, J.; Wiedemann, H.; Glanville, R.W.; Timpl, R.
 Eur. J. Biochem. 139, 401-410, 1984
 A:Title: Subunit structure and assembly of the globular domain of basement-membrane collagen
 A:Reference number: S17801; MUID:84132058
 A:Accession: S19086
 A:Molecule type: protein
 A:Residues: 1475-1481, 'X', 1483-1487 <WEB>
 C:Genetics:
 A:Introns: 15/2; 33/3; 963/1; 1003/3; 1064/3; 1085/3
 C:Superfamily: collagen alpha 1(IV) chain
 C:Keywords: Basement membrane; cell binding; coll; coll; disulfide bond; duplication; e
 F:1-28/Domain: signal sequence status predicted <SIG>
 F:29-1707/Product: collagen alpha 1(IV) chain #status predicted <MAT>
 F:29-171/Domain: 75 #status predicted <7SD>
 F:58-1480/Domain: collagenous #status predicted <COL>
 F:141-143/Region: cell attachment (R-G-D) motif
 F:360-362/Region: cell attachment (R-G-D) motif
 F:779-781/Region: cell attachment (R-G-D) motif
 F:884-886/Region: cell attachment (R-G-D) motif
 F:965-967/Region: cell attachment (R-G-D) motif
 F:1223-1225/Region: cell attachment (R-G-D) motif
 F:1447-1449/Region: cell attachment (R-G-D) motif
 F:1481-1707/Domain: carboxyl-terminal nonhelical, NCI #status predicted <NCI>
 F:1589-1589/Domain: repeat NCI #status predicted <NCI1>
 F:1590-1707/Domain: repeat NCI #status predicted <NCI2>
 F:42,47,51,53,481,483/Disulfide bonds: interchain #status predicted
 F:138,1270/Binding site: carbohydrate bonds: (Asn) (covalent) #status predicted
 F:656-676,1544-1550,1650-1660/Disulfide bonds: #status predicted
 F:985-988,997,1003,1028,1031,1067,1101,1113,1119,1143,1170,1200,1242,1305,1368,1391/Bind
 F:985,988,997,1003,1028,1031,1067,1101,1113,1119,1143,1170,1200,1242,1305,1368,1391/Mod
 F:1009,1012,1018,1021,1024,1037,1040,1043,1046,1052,1058,1070,1098,1110,1128,1140,1149,1
 77,1383,1386,1401,1408,1420,1423,1429,1444,1465,1468,1471,1477/Modified site: hydroxypro

Query Match 70.7%; Score 65; DB 2; Length 1707;
 Best Local Similarity 66.7%; Pred. No. 0.12;
 Matches 10; Conservative 2; Mismatches 3; Indels 0; Gaps 0;
 Oy 1 GVGKDKGNPGWPGAP 15
 1:|||||11111
 Db 1165 GIPGDKGDFGWEPEP 1179

RESULT 7
 CGH02B
 collagen alpha 2(IV) chain precursor - human
 N:Alternate names: procollagen alpha 2(IV) chain
 C:Species: Homo sapiens (man)
 C>Date: 07-Jun-1990 #sequence-revision 03-Oct-1995 #text-change 22-Jun-1999
 C:Accession: A32024; S00007; S00246; S1678; S16911; B32117; S16877; S00165; S39
 R:Hostikra, S.L.; Tryggvason, K.
 J. Biol. Chem. 263, 19488-19493, 1988
 A:Title: The complete primary structure of the alpha2 chain of human type IV collagen an
 A:Reference number: A32024; MUID:89066769
 A:Accession: A32024
 A:Molecule type: mRNA
 A:Residues: 1-1712 <HOS1>
 A:Cross-references: EMBL:J04210; EMBL:X05610; GB:M20753; NID:q29550; PIDN:CAA29098.1; PI
 R:Hostikra, S.L.; Kurkinen, M.; Tryggvason, K.
 FEBS Lett. 216, 281-286, 1987
 A:Title: Nucleotide sequence coding for the human type IV collagen alpha-2 chain cDNA re
 A:Reference number: S00007; MUID:87219158
 A:Accession: S00007
 A:Molecule type: mRNA
 A:Residues: 1254-1398, 'V', 1400-1712 <HOS2>

A:Cross-references: EMBL:J04210; EMBL:X05610; GB:M20753; NID:q29550; PIDN:CAA29098.1;
 A:Note: 1399-116 was also found
 R:Hostikra, S.L.; Tryggvason, K.
 FEBS Lett. 224, 297-305, 1987
 A:Title: Extensive structural differences between genes for the alpha(1) and alpha(2)
 A:Reference number: S02624; MUID:88083553
 A:Accession: S02624
 A:Status: not compared with conceptual translation
 A:Molecule type: DNA
 A:Residues: 1347-1350,1377-1383,1426-1432,1465-1471,1529-1535,1625-1630 <HOS3>
 A:Note: complete nucleotide sequence not shown
 R:Brazel, D.; Pollner, R.; Oberbauer, I.; Kuehn, K.
 Eur. J. Biochem. 172, 35-42, 1988
 A:Title: Human basement membrane collagen (type IV): the amino acid sequence of the a
 A:Reference number: S00246; MUID:88151998
 A:Accession: S00246
 A:Molecule type: mRNA
 A:Residues: 1-682, 'G', 684-1043 <BRA>
 A:Cross-references: EMBL:X05562; NID:q30075; PIDN:CAA29076.1; PID:q30076
 R:Oberbauer, I.
 submitted to the EMBL Data Library, June 1987
 A:Reference number: S17678
 A:Accession: S17678
 A:Molecule type: mRNA
 A:Residues: 1-470, 'P', 472-682, 'G', 684-1043 <OBE>
 A:Cross-references: EMBL:X05562; NID:q30075; PIDN:CAA29076.1; PID:q30076
 R:Poeschl, E.; Pollner, R.; Kuehn, K.
 EMBO J. 7, 2687-2695, 1988
 A:Title: The genes for the alpha(IV) and alpha2(IV) chains of human basement membran
 A:Reference number: S02738; MUID:89030632
 A:Accession: S16911
 A:Status: translation not shown
 A:Molecule type: DNA
 A:Residues: 1-33 <OBE>
 A:Cross-references: EMBL:X12784; GB:M36963; NID:q30072; PIDN:CAA31275.1; PID:q30073
 R:Solinen, R.; Huotari, M.; Hostikra, S.L.; Prockop, D.J.; Tryggvason, K.
 J. Biol. Chem. 263, 17217-17220, 1988
 A:Title: The structural genes for alpha1 and alpha2 chains of human type IV collagen
 A:Reference number: A92690; MUID:89034231
 A:Accession: B32117
 A:Molecule type: DNA
 A:Residues: 1-33 <SO11>
 A:Cross-references: EMBL:J04217; EMBL:J05039; NID:q180759; PIDN:AAAS3097.1; PID:q5532
 R:Solinen, R.; Huotari, M.; Ganguy, A.; Prockop, D.J.; Tryggvason, K.
 J. Biol. Chem. 264, 13565-13571, 1989
 A:Title: Structural organization of the gene for the alpha-1 chain of human type IV c
 A:Reference number: S16876; MUID:89240433
 A:Accession: S16877
 A:Status: nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-33 <SO12>
 A:Cross-references: EMBL:J04217; NID:q180759; PIDN:AAAS3097.1; PID:q5532
 A:Note: this sequence was submitted to the EMBL Data Library, October 1988
 R:Stiebold, B.; Qian, R.O.; Glanville, R.W.; Hofmann, H.; Deutzmann, R.; Kuehn, K.
 Eur. J. Biochem. 168, 569-575, 1987
 A:Title: Construction of a model for the aggregation and cross-linking region (7S dom
 A:Reference number: S00165; MUID:88029476
 A:Accession: S00165
 A:Molecule type: protein
 A:Residues: 37-247 <STF1>
 A:Note: the sequence from Fig. 4 is inconsistent with that from Fig. 3 in having 175-
 R:Eble, J.A.; Golbik, R.; Mann, K.; Kuehn, K.
 EMBO J. 12, 4795-4802, 1993
 A:Title: The alpha-1-beta-1 integrin recognition site of the basement membrane collag
 A:Reference number: S39614; MUID:94038963
 A:Accession: S39615
 A:Molecule type: protein
 A:Residues: 407-570 <EBL>
 R:MacRicht, R.S.; Benson, V.A.; Lovello, K.T.; van der Rest, M.; Fietzek, P.P.
 Biochemistry 22, 4940-4948, 1983
 A:Title: Isolation and characterization of pepsin-solubilized human basement membrane
 A:Reference number: S16910; MUID:84053346

A:Accession: S16912
 A:Molecule type: protein
 A:Residues: 490-492, 'X', 494-496:675-677, 'G', 679-680, 'G', 682, 684-685, 'P' <MAC>
 A:Experimental source: placenta
 R:Clanville, R.W.; Rauter, A.
 Hoppe-Seyler's Z. Physiol. Chem. 362, 943-951, 1991
 A:Title: Peptide fragments of human placental basement-membrane collagens showing interrupted
 A:Reference number: S16908; MUID:82005835
 A:Accession: B58517
 A:Molecule type: protein
 A:Residues: 490-492, 'X', 494-501, 'P', 503-507:952-957, 'X', 959-966, 'X', 968, 984-986, 'X', 988-
 81-1195 <GLA>
 R:Kilien, P.D.; Francomano, C.A.; Yamada, Y.; Modi, W.S.; O'Brien, S.J.
 Hum. Genet. 77, 318-324, 1987
 A:Title: Partial structure of the human alpha-2(IV) collagen chain and chromosomal local
 A:Reference number: S01450; MUID:88085168
 A:Accession: S01450
 A:Molecule type: mRNA
 A:Residues: 1040, 'L', 1042-1398, 'V', 1400-1418, 'M', 1420-1635, 'V', 1637-1712 <KIL>
 A:Cross-references: EMBL:M24766; NID:9537328; PIDN:AAA52043.1; PID:9537329
 R:Siebold, B.; Deutzmann, R.; Kuehn, K.
 Eur. J. Biochem. 176, 617-624, 1988
 A:Title: The arrangement of intra- and intermolecular disulfide bonds in the carboxyterm
 A:Reference number: S02550; MUID:89005112
 A:Accession: S02550
 A:Molecule type: protein
 A:Residues: 1480-1535:1545-1614:1617-1662, 'H', 1664-1700, 'G', 1705-1708:1710-1712 <SIE2>
 A:Note: the sequence form Fig. 7 is inconsistent with that shown in Fig. 11 in having 17
 R:Myers, U.C.; Howard, P.S.; Jelen, A.M.; Dion, A.S.; Macarak, E.J.
 J. Biol. Chem. 262, 9231-9238, 1987
 A:Title: Duplication of type IV collagen COOH-terminal repeats and species-specific expr
 A:Reference number: A27114; MUID:87250571
 A:Accession: B27114
 A:Molecule type: mRNA
 A:Residues: 1486-1574, 'I', 1576-1712 <MYE>
 A:Cross-references: EMBL:J02760; NID:9180425; PIDN:AAA58422.1; PID:9180426
 C:Comment: Prolines and lysines at the third position of the tripeptide repeating unit c
 ed and subsequently O-glycosylated.
 C:Genetics:
 A:Gene: GDB:COL4A2
 A:Cross-references: GDB:119792; OMIM:120090
 A:Map position: 13q24-13q34
 A:Intons: 13/2, 33/3, 134/1, 1380/1, 1429/1, 1468/1, 1532/1, 1527/3 #status incomplete
 A:Note: the alpha 1(IV) and alpha 2(IV) chain genes are encoded on opposite strands with
 C:Complex: Type IV collagen is a heterotrimer of two alpha 1(IV) chains (see PIR:CGH04B)
 domains (with disulfide and desmosine cross-links), dimeric associations among trimer co
 rrupted helical domain (with disulfide and desmosine cross-links).
 C:Function:
 A:Description: structural component of basement membrane
 C:Superfamily: collagen alpha 1(IV) chain
 C:Keywords: basement membrane; cell binding; coll; extracellular matrix; glycopro
 F:1-28/Domain: signal sequence #status predicted <SIG>
 F:29-1712/Product: collagen alpha 2(IV) chain #status predicted <MAT>
 F:29-57/Domain: amino-terminal nonhelical, NH1 <NH1>
 F:58-1485/Region: interrupted helical
 F:58-364/Region: cell attachment (R-G-D) motif
 F:784-786/Region: cell attachment (R-G-D) motif
 F:868-870/Region: cell attachment (R-G-D) motif
 F:889-891/Region: cell attachment (R-G-D) motif
 F:970-972/Region: cell attachment (R-G-D) motif
 F:1069-1071/Region: cell attachment (R-G-D) motif
 F:1228-1230/Region: cell attachment (R-G-D) motif
 F:1452-1454/Region: cell attachment (R-G-D) motif
 F:1486-1712/Domain: carboxyl-terminal nonhelical, NC1 <NC1>
 F:1495-1593/Region: collagen IV carboxyl-terminal repeat <CR1>
 F:1603-1708/Domain: collagen IV carboxyl-terminal repeat <CR2>
 F:42,47,51,53,137,483,485/Disulfide bonds: Interchain #status predicted
 F:57,87,90,102,165,168,225,233,242/Binding site: carbohydrate (lys) (covalent) #status p
 F:57/Modified site: 5-hydroxylysine (lys) #status atypical
 F:53,75,96,114,120,123,132,150,159,186,189,198,201,213,216,219,496,499,955,964,1103,1115
 F:87,90,102,165,168,225,239,242/Modified site: 5-hydroxylysine (lys) #status experimental
 F:138/Binding site: carbohydrate (Asn) (covalent) #status atypical
 F:209/Modified site: 4-hydroxyproline (Pro) #status atypical

F:661-681/Disulfide bonds: #status predicted
 F:1275/Binding site: carbohydrate (Asn) (covalent) #status predicted
 F:1504-1590,1537-1593/Disulfide bonds: (or 1504-1593, 1537-1590) #status experimental
 F:1549-1555,1658-1665/Disulfide bonds: #status experimental
 F:1612-1705,1646-1708/Disulfide bonds: (or 1612-1708, 1646-1705) #status experimental

Query Match 69.6%; Score 64; DB 1; Length 1712;
 Best Local Similarity 73.3%; Pred. No. 0.17;
 Matches 11; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 GVGKDGKNGPMPGAP 15
 Db 1368 GPKGPKDPPGPGAP 1382

RESULT 8
 S23296
 collagen alpha 2(IX) chain precursor - chicken
 C:Species: Gallus gallus (chicken)
 C:Date: 22-Nov-1993 #sequence, revision 09-Mar-1996 #text, change 20-Sep-1999
 C:Accession: S23296; C34493; A29032; S22062; B28754; A18856; B18856; A28518;
 R:Ninomiya, Y.; Castagnola, P.; Gerecke, D.; Gordon, M.K.; Jacenko, O.; Luvall, P.;
 Mauchl, N.; Olsen, B.R.
 in Extracellular Matrix Genes, Sandell L.J. and Boyd C.D., eds., pp.79-114, Academic
 A:Title: The molecular biology of collagens with short triple-helical domains.
 A:Reference number: S22243
 A:Accession: S23296
 A:Status: not compared with conceptual translation
 A:Molecule type: DNA
 A:Residues: 1-677 <N1>
 R:Nishimura, I.; Muragaki, Y.; Olsen, B.R.
 J. Biol. Chem. 264, 20033-20041, 1989
 A:Title: Tissue-specific forms of type IX collagen-proteoglycan arise from the use of
 A:Reference number: A34493; MUID:90062114
 A:Accession: C34493
 A:Molecule type: mRNA
 A:Residues: 1-174 <N1>
 A:Cross-references: EMBL:M28660; NID:9211625; PIDN:AAA48709.1; PID:9211626
 R:McCormick, D.; van der Rest, M.; Goodship, J.; Lozano, G.; Ninomiya, Y.; Olsen, B.R.
 Proc. Natl. Acad. Sci. U.S.A. 84, 4044-4048, 1987
 A:Title: Structure of the glycosaminoglycan domain in the type IX collagen-proteoglyc
 A:Reference number: A29032; MUID:87231947
 A:Accession: A29032
 A:Molecule type: mRNA
 A:Residues: 120-195 <MCC>
 A:Cross-references: GB:M16715; NID:9211365; PIDN:AAA48644.1; PID:9555431
 A:Accession: B29032
 A:Molecule type: protein
 A:Residues: 147-165:170-180, 'X', 182-188 <MC2>
 R:Lozano, G.; Olsen, B.R.
 submitted to the EMBL Data Library, April 1990
 A:Reference number: S22062
 A:Accession: S22062
 A:Molecule type: DNA
 A:Residues: 401, 'RA', 404-631, 'D', 633-677 <LO1>
 A:Cross-references: EMBL:M11327
 R:Lozano, G.; Ninomiya, Y.; Thompson, H.; Olsen, B.R.
 Proc. Natl. Acad. Sci. U.S.A. 82, 4050-4054, 1985
 A:Title: A distinct class of vertebrate collagen genes encodes chicken type IX collag
 A:Reference number: A94705; MUID:85216631
 A:Accession: B28754
 A:Molecule type: DNA
 A:Residues: 404-633:644-677 <LO2>
 A:Cross-references: EMBL:M11327
 A:Note: This sequence has been revised in reference S22062
 R:Ninomiya, Y.; van der Rest, M.; Mayne, R.; Lozano, G.; Olsen, B.R.
 Biochemistry 24, 4223-4229, 1985
 A:Title: Construction and characterization of cDNA encoding the alpha2 chain of chick
 A:Reference number: A18856; MUID:86026268
 A:Accession: A18856
 A:Molecule type: mRNA
 A:Residues: 401, 'RA', 404-677 <N1>

A:Accession: B18856
 A:Molecule type: protein
 A:Residues: 433-435, 'X', 437-450:542-562 <NT3>
 R:Huber, S.; Witterthaler, K.H.; Vaughan, L.
 J. Biol. Chem. 263, 752-756, 1988
 A:Title: Isolation and sequence analysis of the glycosaminoglycan attachment site of type
 A:Reference number: A28518; MUID:88087195
 A:Accession: A28518
 A:Molecule type: protein
 A:Residues: 156-159, 'X', 161-166, 'X', 168-173, 'X', 175-178 <HUB>
 R:van der Rest, M.; Mayne, R.
 J. Biol. Chem. 263, 1615-1618, 1988
 A:Title: Type IX collagen proteoglycan from cartilage is covalently cross-linked to type
 A:Reference number: S23813; MUID:88115274
 A:Accession: S23813
 A:Molecule type: protein
 A:Residues: 170-180, 'X', 182-184 <VAN>
 A:Note: Evidence for aldolase cross-linkage of 190-Lys to collagen alpha 1(II) chain is
 R:Mayne, R.; van der Rest, M.; Nihmly, Y.; Olsen, B.R.
 Ann. N. Y. Acad. Sci. 460, 38-46, 1985
 A:Title: The structure of type IX collagen.
 A:Reference number: S22238; MUID:86185164
 A:Accession: S22238
 A:Molecule type: protein
 A:Residues: 542-567 <MAY>
 C:Genetics:
 A:Introns: 427/3; 439/3; 454/3; 465/3; 514/3; 533/1; 596/1; 622/1
 C:Superfamily: unassigned collagens
 C:Keywords: chondroitin sulfate proteoglycan; collid coll; extracellular matrix; glycop
 F:1-22/Domain: signal sequence #status predicted <SIG>
 F:22-677/Product: collagen alpha 2(XI) chain #status predicted <MAT>
 F:25-161/Domain: collagen COL3 #status predicted <COL3>
 F:162-178/Domain: non-collagenous NC3 #status predicted <NC3>
 F:179-517/Domain: collagenous COL2 #status predicted <COL2>
 F:518-547/Domain: non-collagenous NC2 #status predicted <NC2>
 F:548-662/Domain: collagenous COL1 #status predicted <COL1>
 F:663-677/Domain: non-collagenous NC1 #status predicted <NC1>
 F:158-178/Modified site: 4-hydroxyproline (Pro) #status experimental
 F:167/Binding site: chondroitin sulfate (Ser) (covalent) #status experimental
 F:181/Modified site: 5-hydroxylysine (Lys) #status experimental
 F:181/Binding site: carboxylate (Lys) (covalent) #status experimental
 F:190/Modified site: allysine (Lys) #status predicted

Query Match 68.5%; Score 63; DB 2; Length 677;
 Best Local Similarity 66.7%; Pred. No. 0.094;
 Matches 10; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

OY 1 GYKGRKNGPWCAP 15
 I: ||||| |||||
 DB 46 GIDGDKGSPGAPSP 60

RESULT 9
 A:Accession: A54849
 A:Molecule type: mRNA
 A:Residues: 1-294 <CHR>
 A:Cross-references: GB:I02870; NID:987124; PIDN:AA75438.1; PID:9987125
 R:Tanaka, T.; Takahashi, K.; Furukawa, F.; Imamura, S.
 Biochem. Biophys. Res. Commun. 183, 958-963, 1992
 A:Title: Molecular cloning and characterization of type VII collagen cDNA.
 A:Reference number: PH0844; MUID:92231902

A:Accession: PH0844
 A:Molecule type: mRNA
 A:Residues: 'EPR', 340-475, 'BALSTASHSTICMRATRMWPCNRGSHWTRACPCNRPASRRAARAG', 524-528, '
 A:Cross-references: DDBJ:D11152; DDBJ:D13694; NID:9453658; PIDN:BA02853.1; PID:94536
 A:Experimental source: keratinocyte
 A:Note: The authors translated the codon ACC for residues 394 and 397 as Tyr
 R:Parente, M.G.; Chung, L.C.; Ryyanen, J.; Woodley, D.T.; Wynn, K.C.; Bauer, E.A.;
 Proc. Natl. Acad. Sci. U.S.A. 88, 6931-6935, 1991
 A:Title: Human type VII collagen: cDNA cloning and chromosomal mapping of the gene.
 A:Reference number: S16316; MUID:91334380
 A:Accession: S16316
 A:Molecule type: mRNA
 A:Residues: 815-892, 'E', 894-1439 <PAR>
 A:Cross-references: GB:M65158; GB:S49017; NID:9180914; PIDN:AA96439.1; PID:9180915
 A:Experimental source: keratinocyte
 R:Gannon, W.R.; Abernethy, M.L.; Padilla, K.M.; Prisyah, P.S.; Cook, M.E.; Wright,
 J. Invest. Dermatol. 99, 691-696, 1992
 A:Title: Noncollagenous (NC1) domain of collagen VII resembles multidomain adhesion p
 A:Reference number: I56328; MUID:93107742
 A:Accession: I56328
 A:Molecule type: protein
 A:Status: translated from GB/EMBL/DBJ
 A:Residues: 'EPR', 372-517, 'DV', 520-540, 'W', 542-1255 <RES>
 A:Cross-references: GB:S51236; NID:9262308; PIDN:AA824637.1; PID:9262309
 R:Selzler, J.L.; Eisen, A.Z.; Bauer, E.A.; Morris, N.P.; Glanville, R.W.; Burgeson, R
 J. Biol. Chem. 264, 3822-3826, 1989
 A:Title: Cleavage of type VII collagen by interstitial collagenase and type IV collag
 A:Reference number: A30296; MUID:89139437
 A:Accession: A30296
 A:Molecule type: protein
 A:Residues: 'A', 1240-1246, 'G', 1248-1250, 'XE', 1253-1255, 'Q', 1257, 'E', 2032, 'C', 2034-204
 A:Note: two reported peptides cannot be reliably located
 R:Greenspan, D.S.
 Hum. Mol. Genet. 2, 273-278, 1993
 A:Title: The carboxyl-terminal half of type VII collagen, including the non-collagene
 A:Reference number: I48103; MUID:93271985
 A:Accession: I48103
 A:Molecule type: mRNA
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Residues: 2395-2871, 'S', 2873-2944 <RE2>
 A:Cross-references: GB:I06862; NID:9388713; PIDN:AA89196.1; PID:9388714
 R:Christiano, A.M.; Ryyanen, M.; Uitto, J.
 Proc. Natl. Acad. Sci. U.S.A. 91, 3549-3553, 1994
 A:Title: Dominant dystrophic epidermolysis bullosa: identification of a Gly --> Ser s
 A:Reference number: A55255; MUID:94224777
 A:Contents: annotation
 C:Comment: Prolines and lysines at the third position of the tripeptide repeating uni
 ed and subsequently O-glycosylated.
 C:Genetics:
 A:Gene: GDB:COL7A1; EBR1; EBD1; EB
 A:Cross-references: GDB:128750; OMIM:120120
 A:Map position: 3p21.3-3p21.3
 A:Note: defects in this gene can result in dominant and recessive dystrophic epidermo
 C:Function:
 C:Complex: type VII collagen is probably a homotrimer
 A:Description: structural component of extracellular polymer associated with anchurin
 C:Superfamily: unassigned collagens; animal knitz-type proteinase inhibitor homology
 C:Keywords: collid coll; extracellular matrix; glycoprotein; hydroxylysine; hydroxypr
 F:1-16/Domain: signal sequence #status predicted <SIG>
 F:17-2944/Product: collagen alpha 1(VII) chain #status predicted <MAT>
 F:17-1253/Domain: amino-terminal nonhelical #status predicted <NC1>
 F:36-201/Domain: von Willebrand factor type A repeat homology <VWA1>
 F:231-318/Domain: fibronectin type III repeat homology <FN1>
 F:327-413/Domain: fibronectin type III repeat homology <FN2>
 F:414-502/Domain: fibronectin type III repeat homology <FN3>
 F:508-593/Domain: fibronectin type III repeat homology <FN4>
 F:598-683/Domain: fibronectin type III repeat homology <FN5>
 F:686-771/Domain: fibronectin type III repeat homology <FN6>
 F:776-863/Domain: fibronectin type III repeat homology <FN7>
 F:864-952/Domain: fibronectin type III repeat homology <FN8>
 F:954-1045/Domain: fibronectin type III repeat homology <FN9>
 F:1052-1219/Domain: von Willebrand factor type A repeat homology <VWA2>

N:Alternate names: procollagen alpha 4(IV) chain
 C:Species: Homo sapiens (man)
 C>Date: 06-Feb-1995 #sequence_rev13on 03-Oct-1995 #text_change 16-Jun-2000
 C:Accession: A55360; S36854; S28777
 R:Reinonen, A.; Maruyama, M.; Mochizuki, T.; Tryggvason, K.; Reenders, S.T.
 J. Biol. Chem. 269, 26172-26177, 1994
 A:Title: Complete primary structure of the human type IV collagen alpha4(IV) chain. Comp
 A:Reference number: A55360; MUID:95014445
 A:Accession: A55360
 A:Status: nucleic acid sequence not shown
 A:Molecule type: mRNA
 A:Residues: 1-1690 <EB>
 A:Cross-references: GB:X81053; NID:9574805; PIDN:CAA56943.1; PID:9574806
 R:Sugimoto, M.; Ohashi, T.; Yoshioke, H.; Matsuo, N.; Ninomiya, Y.
 FEBS Lett. 330, 122-128, 1993
 A:Title: cDNA isolation and partial gene structure of the human alpha-4(IV) collagen ch
 A:Reference number: S36854; MUID:93374047
 A:Accession: S36854
 A:Molecule type: DNA
 A:Residues: 1219-1658, 'FE', 1661-1690 <SUG>
 A:Cross-references: DBJ:017391; NID:9440365; PIDN:BA04214.1; PID:9457161
 A:Experimental source: Whole eye
 R:Kamagata, Y.; Matzel, M.G.; Ninomiya, Y.
 J. Biol. Chem. 267, 23753-23758, 1992
 A:Title: Isolation and sequencing of cDNAs and genomic DNAs encoding the alpha4 chain of
 A:Reference number: S28777; MUID:93054733
 A:Accession: S28777
 A:Molecule type: DNA
 A:Residues: 1407-1424, 'G', 1426-1430, 'A', 1432-1439, 'L', 1441-1507 <KAM>
 A:Cross-references: GB:L01475; GB:L01476
 A:Note: the codons given for 1438-Asp (GAG) and 1443-Gly (GCA) are inconsistent with the
 C:Comment: Prolines and lysines at the third position of the tripeptide repeating unit (C
 ed and subsequently O-glycosylated.
 C:Genetics:
 A:Gene: GDB:COL4A4
 A:Cross-references: GDB:132673; OMIM:120131
 A:Map position: 2q35-2q37
 A:Introns: 39/1; 1406/1; 1445/1; 1508/1; 1603/3 #status: incomplete
 A:Note: the alpha 3(IV) and alpha 4(IV) chain genes are encoded on opposite strands with
 C:Complex: this minor type IV collagen is thought to form a heterotrimer of two alpha 3
 er associations in the interrupted helical domain (with disulfide and desmosine cross-l
 C:Function:
 A:Description: minor structural component of extracellular basement membrane in kidney
 C:Superfamily: collagen alpha 1(IV) chain
 C:Keywords: basement membrane; coiled coil; extracellular matrix; glycoprotein; hydroxy
 F:1-38/Domain: signal sequence #status predicted <SIG>
 F:39-1690/Product: collagen alpha 4(IV) chain #status predicted <MAT>
 F:39-61/Domain: amino-terminal nonhelical, NHI <NHI>
 F:62-1466/Region: interrupted helical
 F:94-96/Region: cell attachment (R-G-D) motif
 F:145-147/Region: cell attachment (R-G-D) motif
 F:189-191/Region: cell attachment (R-G-D) motif
 F:310-312/Region: cell attachment (R-G-D) motif
 F:724-726/Region: cell attachment (R-G-D) motif
 F:789-787/Region: cell attachment (R-G-D) motif
 F:989-991/Region: cell attachment (R-G-D) motif
 F:1212-1214/Region: cell attachment (R-G-D) motif
 F:1467-1690/Domain: carboxyl-terminal nonhelical, NCI <NCI>
 F:1471-1569/Domain: collagen IV carboxyl-terminal repeat <CT1>
 F:1579-1666/Domain: collagen IV carboxyl-terminal repeat <CT2>
 F:7, 52, 55, 57, 266, 400, 492, 494, 666, 790, 828, 1095, 1131, 1294, 1317, 1375, 1407/Disulfide bc
 F:142, 669/Binding site: carbohydrate (Asn) (covalent) #status predicted
 F:1480-1563, 1513-1569/Disulfide bonds: (or 1480-1569, 1513-1566) #status predicted
 F:1525-1531, 1634-1641/Disulfide bonds: #status predicted
 F:1588-1683, 1622-1686/Disulfide bonds: (or 1588-1686, 1622-1683) #status predicted

Query Match 66.3%; Score 61; DB 1; Length 1690;
 Best Local Similarity 60.0%; Pred. No. 0.46;
 Matches 9; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 GVKGDKGNPGMPGAP 15

DB 184 GIOGDRDGPGLPGLP 198
 RESULT 16
 S16366
 collagen alpha 2(IV) chain precursor - pig roundworm
 C:Species: Ascaris suum (pig roundworm)
 C>Date: 04-Dec-1992 #sequence_rev13on 04-Dec-1992 #text_change 13-Aug-1999
 C:Accession: S16366
 R:Pettilt, J.; Kingston, I.B.
 J. Biol. Chem. 266, 16149-16156, 1991
 A:Title: The complete primary structure of a nematode alpha-2(IV) collagen and the pa
 A:Reference number: S16366; MUID:91340768
 A:Accession: S16366
 A:Molecule type: mRNA
 A:Residues: 1-1763 <JB>
 A:Cross-references: GB:M67507; NID:9159648; PIDN:AAA18014.1; PID:9159649
 C:Genetics:
 A:Introns: 229/3; 266/3; 305/3; 360/3; 424/1; 489/1; 548/1; 656/3; 790/1; 891/1; 963/
 C:Superfamily: collagen alpha 1(IV) chain
 C:Keywords: alternative splicing; basement membrane; cell binding; coiled coil; disul
 F:1-26/Domain: signal sequence #status predicted <SIG>
 F:27-1763/Product: collagen alpha 2(IV) chain #status predicted <MAT>
 F:27-42/Domain: non-collagenous NHI #status predicted <NHI>
 F:43-1529/Domain: collagenous #status predicted <COL>
 F:197-199/Region: cell attachment (R-G-D) motif
 F:1530-1763/Domain: carboxyl-terminal nonhelical, NCI #status predicted <NCI>
 F:1530-1638/Domain: repeat NCI #status predicted <NCI2>
 F:1639-1763/Domain: repeat NCI #status predicted <NCI2>
 F:31, 34, 39, 41, 536, 539/Disulfide bonds: interchain #status predicted
 F:126/Binding site: carbohydrate (Asn) (covalent) #status predicted
 F:1593-1599, 1702-1709/Disulfide bonds: #status predicted

Query Match 66.3%; Score 61; DB 2; Length 1763;
 Best Local Similarity 73.3%; Pred. No. 0.48;
 Matches 11; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 GVKGDKGNPGMPGAP 15
 DB 760 GLPGMKGNPGLPGLP 774

RESULT 17
 S10696
 collagen alpha 1(III) chain - chicken (fragment)
 C:Species: Gallus gallus (chicken)
 C>Date: 13-Sep-1996 #sequence_rev13on 13-Sep-1996 #text_change 13-Aug-1999
 C:Accession: I50696
 R:Neh, H.D.; Niu, Z.; Adams, S.L.
 J. Biol. Chem. 269, 16443-16448, 1994
 A:Title: An alternative transcript of the chick type III collagen gene that does not
 A:Reference number: A54041; MUID:94266842
 A:Accession: I50696
 A:Status: Preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-310 <NAH>
 A:Cross-references: EMBL:U07974; NID:9520456; PIDN:AAA83409.1; PID:9537433
 C:Genetics:
 A:Gene: COL3A1
 C:Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homolo

Query Match 65.2%; Score 60; DB 2; Length 310;
 Best Local Similarity 71.4%; Pred. No. 0.12;
 Matches 10; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 GVKGDKGNPGMPGA 14
 DB 281 GVKGERGSPGPGCA 294

RESULT 18

A45137

collagen alpha 4(IV) chain - rabbit

C:Species: Oryctolagus cuniculus (domestic rabbit)

C:Date: 30-Apr-1993 #sequence_revision 18-Nov-1994 #text_change 17-Mar-1999

C:Accession: A45137

R:Kamagata, Y.; Mattei, M.G.; Nishimura, Y.

J. Biol. Chem. 267, 23753-23758, 1992

A:Title: Isolation and sequencing of cDNAs and genomic DNAs encoding the alpha4 chain of

A:Reference number: S28777; MUID:93054733

A:Accession: A45137

A:Status: preliminary; not compared with conceptual translation

A:Molecule type: mRNA

A:Residues: 1-623 <KAM>

A:Experimental source: basement membrane

A>Note: sequence extracted from NCBI backbone (NCBI:118549)

C:Superfamily: collagen alpha 1(IV) chain

Query Match

Best Local Similarity 65.2%; Score 60; DB 2; Length 623;*

Matches 10; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 GVKGDKGNPGPG 13

I:|||||I|I|I|

DB 139 GIKGDKGEPGSPG 151

RESULT 19

S41067

collagen alpha 1(III) chain - rat

C:Species: Rattus norvegicus (Norway rat)

C:Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 13-Aug-1999

C:Accession: S41067; A29905; S31924

R:Glimoff, V.; Maekelae, J.K.; Vuorio, E.

Biochim. Biophys. Acta 1217, 41-48, 1994

A:Title: Cloning of cDNA for rat pro alpha-1(III) collagen mRNA. Different expression pa

A:Reference number: S41067; MUID:94114571

A:Accession: S41067

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-636 <GLD>

A:Cross-references: EMBL:X70369; NID:957915; PIDN:CAA49832.1; PID:957916

R:Pranke, F.R.; Hsu, C.Y.J.; Myers, J.C.; Lin, E.; Lytle, C.R.; Komm, B.; Mohn, K.

DNA 7, 347-354, 1988

A:Title: Regulation of alpha-2(I), alpha-1(III), and alpha-2(V) collagen mRNAs by estr

A:Reference number: A29905; MUID:88296083

A:Accession: A29905

A:Molecule type: mRNA

A:Residues: 308-482 <FRA>

A:Cross-references: GB:M21354; NID:9203500; PIDN:AAA40942.1; PID:9203501

R:Glimoff, V.; Maekelae, J.K.; Vuorio, E.

submitted to the EMBL Data Library, February 1993

A:Reference number: S31924

A:Accession: S31924

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 2-636 <GLD>

A:Cross-references: EMBL:X70369

C:Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;

C:Keywords: coiled coil; extracellular matrix; glycoprotein; trimer; triple helix

F:408-636/Domain: fibrillar collagen carboxyl-terminal homology <PCC>

Query Match

Best Local Similarity 71.4%; Score 60; DB 2; Length 636;

Matches 10; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 GVKGDKGNPGPG 14

I:|||||I|I|I|

DB 30 GVKGDKGEPGPG 43

RESULT 20

I50694

collagen alpha 1(III) chain - chicken (fragment)

C:Species: Gallus gallus (chicken)

C:Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 13-Aug-1999

C:Accession: I50694

R:Nah, H.D.; Niu, Z.; Adams, S.L.

J. Biol. Chem. 269, 16443-16448, 1994

A:Title: An alternative transcript of the chick type III collagen gene that does not

A:Reference number: A54041; MUID:94266842

A:Accession: I50694

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-886 <NAH>

A:Cross-references: EMBL:U07973; NID:9520454; PIDN:AAA83407.1; PID:9537432

C:Genetics:

A:Gene: COL3A1

C:Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homolo

F:30-90/Domain: von Willebrand factor type C repeat homology <VMC>

Query Match

Best Local Similarity 71.4%; Score 60; DB 2; Length 886;

Matches 10; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 GVKGDKGNPGPG 14

I:|||||I|I|I|

DB 857 GVKGDKGEPGPG 870

RESULT 21

CGB075

collagen alpha 1(III) chain - bovine

C:Species: Bos primigenius taurus (cattle)

C:Date: 04-Dec-1986 #sequence_revision 04-Dec-1986 #text_change 07-May-1999

C:Accession: A02862; A38001; A38002; A38003; A38004; A38005; S71946

R:Fietzek, P.P.; Allmann, H.; Rautenberg, U.; Henkel, W.; Wachter, E.; Kuehn, K.

Hoppe-Seyler's Z. Physiol. Chem. 360, 809-820, 1979

A:Title: The covalent structure of calf skin type III collagen. I. The amino acid seq

A:Reference number: A02862; MUID:80026026

A:Accession: A02862

A:Molecule type: protein

A:Residues: 1-242 <FIE>

R:Dewes, H.; Fietzek, P.P.; Kuehn, K.

Hoppe-Seyler's Z. Physiol. Chem. 360, 821-832, 1979

A:Title: The covalent structure of calf skin type III collagen. II. The amino acid se

A:Reference number: A38001; MUID:80026027

A:Accession: A38001

A:Molecule type: protein

A:Residues: 243-422 <DEW1>

R:Bentz, H.; Fietzek, P.P.; Kuehn, K.

Hoppe-Seyler's Z. Physiol. Chem. 360, 833-840, 1979

A:Title: The covalent structure of calf skin type III collagen. III. The amino acid s

A:Reference number: A38002; MUID:80026028

A:Accession: A38002

A:Molecule type: protein

A:Residues: 423-571 <BEN>

R:Land, H.; Glanville, R.W.; Fietzek, P.P.; Kuehn, K.

Hoppe-Seyler's Z. Physiol. Chem. 360, 841-850, 1979

A:Title: The covalent structure of calf skin type III collagen. IV. The amino acid se

A:Reference number: A38003; MUID:80026029

A:Accession: A38003

A:Molecule type: protein

A:Residues: 572-808 <LAN>

R:Dewes, H.; Fietzek, P.P.; Kuehn, K.

Hoppe-Seyler's Z. Physiol. Chem. 360, 851-860, 1979

A:Title: The covalent structure of calf skin type III collagen. V. The amino acid seq

A:Reference number: A38004; MUID:80026030

A:Accession: A38004

A:Molecule type: protein

A:Residues: 809-947 <DEW2>

R:Allmann, H.; Fietzek, P.P.; Glanville, R.W.; Kuehn, K.

Hoppe-Seyler's Z. Physiol. Chem. 360, 861-868, 1979

A:Title: The covalent structure of calf skin type III collagen. VI. The amino acid sequence.
 A:Reference number: A38005; MUID:80026031
 A:Accession: A38005
 A:Molecule type: protein
 A:Residues: 948-1049 <ALL>
 A:Experimental source: skin
 R:Henkel, W.
 Biochem. J. 318, 497-503, 1996
 A:Title: Cross-link analysis of the C-telopeptide domain from type III collagen.
 A:Reference number: S71946; MUID:96404897
 A:Accession: S71946
 A:Molecule type: protein
 A:Residues: 87-106;1017-1029;1037-1049 <HEN>
 C:Comment: Prolines at the third position of the tripeptide repeating unit (G-X-Y) are H
 C:Comment: The type III collagen molecule is a trimer of identical chains, linked to each
 C:Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;
 C:Keywords: coiled coil; extracellular matrix; glycoprotein; hydroxylysine; hydroxyproline
 F:1-1049/Product: collagen alpha 1(III) chain #status experimental <CAB>
 F:1-14/Region: amino-terminal nonhelical telopeptide
 F:15-1040/Region: helical
 F:587-589/Region: cell attachment (R-G-D) motif
 F:752-754/Region: cell attachment (R-G-D) motif
 F:875-877/Region: cell attachment (R-G-D) motif
 F:878-880/Region: cell attachment (R-G-D) motif
 F:935-937/Region: cell attachment (R-G-D) motif
 F:1041-1049/Region: carboxyl-terminal nonhelical telopeptide
 F:95,107,119,938,950/Modified site: 5-hydroxylysine (Lys) #status experimental
 F:107,950/Modified site: lysine (Lys) #status experimental
 F:107/Binding site: carbohydrate (Lys) (covalent) #status experimental
 F:1040,1041/Dsulfide bonds: Interchain #status predicted

Query Match 65.2%; Score 60; DB 1; Length 1049;
 Best Local Similarity 71.4%; Pred. No. 0.41;
 Matches 10; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

OY 1 GVKGDKNGPCPGA 14
 II:::IIII
 Db 702 GVKGERGSPGPGA 715

RESULT 22
 CGH07L
 collagen alpha 1(III) chain precursor - human
 N:Alternate names: procollagen alpha 1(III) chain
 C:Species: Homo sapiens (man)
 C:Date: 24-Apr-1984 #sequence revision 01-Sep-1995 #text change 21-Jul-2000
 C:Accession: S05272; S04642; PE0011; S01726; S04887; A90399; A94562; I51868; S59511; A90
 R:Prockop, D.J.
 submitted to the EMBL Data Library, February 1989
 A:Reference number: S05272
 A:Accession: S05272
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-1240, V', 1242-1466 <PRC>
 A:Cross-references: EMBL:X14420; NID:930057; PIDN:CAA32583.1; PID:930058
 R:Ala-Kokko, L.; Kontusari, S.; Baldwin, C.T.; Kuivaniemi, H.; Prockop, D.J.
 Biochem. J. 260, 509-516, 1989
 A:Title: Structure of cDNA clones coding for the entire prepro-alpha1(III) chain of human
 erences.
 A:Reference number: S04642; MUID:89350838
 A:Accession: S04642
 A:Molecule type: mRNA
 A:Residues: 1-1196 <ALA>
 A:Cross-references: EMBL:X14420; NID:930057; PIDN:CAA32583.1; PID:930058
 R:Benson-Chanda, V.; Su, M.W.; Well, D.; Chu, M.L.; Ramirez, F.
 Gene 78, 253-265, 1989
 A:Title: Cloning and analysis of the 5' portion of the human type-III procollagen gene
 A:Reference number: PE0011; MUID:89378752
 A:Accession: PE0011
 A:Molecule type: DNA
 A:Residues: 1-176 <BEN>

A:Cross-references: GB:M26539; NID:9180813; PIDN:AAAS2040.1; PID:9180814
 R:Toman, P.D.; Ricca, G.A.; de Crombrughe, B.
 Nucleic Acids Res. 16, 7201, 1988
 A:Title: Nucleotide sequence of a cDNA coding for the amino-terminal region of human
 A:Reference number: S01726; MUID:88303360
 A:Accession: S01726
 A:Molecule type: mRNA
 A:Residues: 1-170 <TOM>
 A:Cross-references: EMBL:X07240; NID:930060; PIDN:CAA30229.1; PID:930061
 A:Note: the authors translated the codon CAG for residue 154 as His
 R:Janeczko, R.A.; Ramirez, F.
 Nucleic Acids Res. 17, 6742, 1989
 A:Title: Nucleotide and amino acid sequences of the entire human alpha-1 (III) collagen
 A:Reference number: S04887; MUID:89386015
 A:Accession: S04887
 A:Molecule type: mRNA
 A:Residues: 149-163, 'G', 164-240, 'D', 242-471, 'D', 473-487, 'L', 489, 'S', 491-613, 'Y', 615-6
 A:Cross-references: EMBL:X15332; NID:929543; PIDN:CAA33387.1; PID:930045
 A:Note: the authors' translation of residues 905-932 is inconsistent with the nucleot
 R:Sever, J.M.; Kang, A.H.
 Biochemistry 16, 1158-1164, 1977
 A:Title: Covalent structure of collagen: amino acid sequence of cyanogen bromide pept
 A:Reference number: A90399; MUID:77134724
 A:Accession: A90399
 A:Molecule type: protein
 A:Residues: 'V', 169-225, 229-277, 'A', 279-292, 'D', 294-398 <SEY1>
 A:Experimental source: liver
 A:Note: sequence corrected by A94562; attachment of 2-O-alpha-D-glucosyl-O-beta-D-gal
 R:Sever, J.M.
 submitted to the Atlas, December 1977
 A:Reference number: A94562
 A:Accession: A94562
 A:Molecule type: protein
 A:Residues: 'V', 169-225, 229-277, 'A', 279-292, 'D', 294-398 <SEY2>
 A:Experimental source: liver
 A:Note: author submitted corrections to A90399
 R:Miliewicz, D.M.; Wiltz, A.M.; Smith, A.C.; Manchester, D.K.; Waldstein, G.; Byers, P.
 Am. J. Hum. Genet. 53, 62-70, 1993
 A:Title: Parental somatic and germ-line mosaicism for a multilexon deletion with unusu
 sporling.
 A:Reference number: I51868; MUID:93304430
 A:Accession: I51868
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 186-194 <ML>
 A:Cross-references: GB:S62925; NID:9386425; PIDN:AAD13937.1; PID:94261637
 R:Chiodo, A.A.; Sillence, D.O.; Cole, W.G.; Bateman, J.F.
 Biochem. J. 311, 939-943, 1995
 A:Title: Abnormal type III collagen produced by an exon-17-skipping mutation of the C
 A:Reference number: S59511; MUID:96067614
 A:Accession: S59511
 A:Molecule type: mRNA
 A:Residues: 302-423 <CHR>
 A:Cross-references: GB:S79677; NID:91195576; PIDN:AAB35615.1; PID:91195577
 R:Sever, J.M.; Kang, A.H.
 Biochemistry 17, 3404-3411, 1978
 A:Title: Covalent structure of collagen: amino acid sequence of five consecutive CNB
 A:Reference number: A90414; MUID:79000343
 A:Accession: A90414
 A:Molecule type: protein
 A:Residues: 399-675, 'N', 677-727 <SEY3>
 A:Experimental source: liver
 R:Lee, B.; Vitale, E.; Superti-Furga, A.; Steimann, B.; Ramirez, F.
 J. Biol. Chem. 266, 5256-5259, 1991
 A:Title: G to T transversion at position +5 of a splice donor site causes skipping of
 A:Reference number: I55349; MUID:91161621
 A:Accession: I55349
 A:Status: translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 537-605 <LEE>
 A:Cross-references: GB:M59312; NID:9180815; PIDN:AAAS2041.1; PID:9180816
 R:Sever, J.M.; Mainardi, C.; Kang, A.H.
 Biochemistry 19, 1583-1589, 1980

A>Title: Covalent structure of collagen: amino acid sequence of alpha1 (III)-CB5 from ty
 A:Reference number: A90438; MUID:801898282
 A:Accession: A90438
 A:Molecule type: protein
 A:Residues: 728-895, 'A', 897-964 <SEX4>
 A:Experimental source: liver
 R:Colo, W.G.; Chlodo, A.A.; Lamande, S.R.; Janeczko, R.; Ramirez, F.; Dahl, H.H.M.; Char
 J: Biol. Chem. 265, 17070-17077, 1990
 A>Title: A base substitution at a splice site in the COL3A1 gene causes exon skipping an
 A:Reference number: A38303; MUID:91009133
 A:Accession: A38303
 A:Molecule type: mRNA
 A:Residues: 861-1015 <COL>
 A:Cross-references: GB:J05617; GB:M55603; GB:M59227; NID:9180678; PIDN:AA59383.1; PID:9
 A:Note: a mutant sequence with 942-977 spliced out from a patient with Ehlers-Danlos syn
 R:ManKo, B.S.; Dalgleish, R.
 Nucleic Acids Res. 16, 2337, 1988
 A>Title: Human pro alpha1(III) collagen: cDNA sequence for the 3' end.
 A:Reference number: S02119; MUID:88189827
 A:Accession: S02119
 A:Molecule type: mRNA
 A:Status: translation not shown
 A:Residues: 950-1018, 'Y', 1020-1183, 'S', 1185-1466 <MAN>
 A:Cross-references: EMBL:X06700; NID:930053; PIDN:CAA29886.1; PID:930054
 R:Sey, J.M.; Kang, A.H.
 Biochemistry 20, 2621-2627, 1981
 A>Title: Covalent structure of collagen: amino acid sequence of alpha1 (III)-CB9 from ty
 A:Reference number: A90446; MUID:81208139
 A:Accession: A90446
 A:Molecule type: protein
 A:Residues: 965-979, 'A', 981-984, 'PS', 987, 'ON', 990-1096, 'P', 1098-1152, 'AR', 1155, 'S', 1157-
 A:Experimental source: liver
 R:Loidl, H.R.; Brinker, J.M.; May, M.; Pihlajaniemi, T.; Morrow, S.; Rosenbloom, J.; Mye
 Nucleic Acids Res. 12, 9383-9394, 1984
 A>Title: Molecular cloning and carboxyl-propeptide analysis of human type III procollage
 A:Reference number: A93551; MUID:85087944
 A:Accession: A93551
 A:Molecule type: mRNA
 A:Residues: 1065-1155, 'P', 1157-1466 <LOI>
 A:Cross-references: EMBL:X01655; EMBL:X01742; NID:929584; PIDN:CAA25821.1
 R:Miskulin, M.; Dalgleish, R.; Klue-Beckerman, B.; Renard, S.I.; Tolstoshev, P.; Brant
 Biochemistry 25, 1408-1413, 1986
 A>Title: Human type III collagen gene expression is coordinately modulated with the type
 A:Reference number: I52393; MUID:86187804
 A:Accession: I52393
 A:Status: translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1161-1200 <MIS>
 A:Cross-references: GB:M13146; NID:9180415; PIDN:AA52003.1; PID:9180416
 R:Emanuel, B.S.; Cannizzaro, L.A.; Seyer, J.M.; Myers, J.C.
 Proc. Natl. Acad. Sci. U.S.A. 82, 3385-3389, 1985
 A>Title: Human alpha 1(III) and alpha 2(V) procollagen genes are located on the long arm
 A:Reference number: I59025; MUID:85216505
 A:Accession: I59025
 A:Status: translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1165-1196 <ENA>
 A:Cross-references: GB:M11134; NID:9180417; PIDN:AA52004.1; PID:9180418
 R:Chu, M.L.; Well, D.; de Wel, W.; Bernard, M.; Sippola, M.; Ramirez, F.
 J. Biol. Chem. 260, 4357-4363, 1985
 A>Title: Isolation of cDNA and genomic clones encoding human pro-alpha1(III) collagen. F
 A:Reference number: A92516; MUID:85157600
 A:Accession: A92516
 A:Molecule type: DNA
 A:Residues: 1176-1240, 'V', 1242-1356, 'P', 1358-1466 <CHU>
 A:Cross-references: GB:M10615; GB:M10793; GB:M10794; GB:M10795; GB:M10796; GB:M10797; GB
 A:Experimental source: liver
 A:Note: the authors translated the codon TTC for residue 1057 as Tyr; the codons given f
 action
 C:Comment: Prolines and lysines at the third position of the tripeptide repeating unit
 3-hydroxylated. About 15% of the lysines are 5-hydroxylated and some are subsequently C
 C:Genetics: GDB:COL3A1
 A:Gene: GDB:COL3A1

A:Cross-references: GDB:118729; OMIM:120180
 A:Map position: 2q31-2q31
 A:Introns: 27/1; 94/3; 149/3; 176/3; 554/3; 587/3; 1175/3; 1275/1; 1337/3; 141
 A:Note: the list of introns is incomplete; defects in this gene can result in Ehlers-
 C:Complex: type III collagen is a homotrimer of monomers initially linked by disulfid
 er of their length, is formed with desmosine cross-links made from lysine and allysin
 C:Function:
 A:Description: structural component of extracellular fibrous polymer that maintains
 C:Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homolo
 C:Keywords: collod coll, Ehlers-Danlos syndrome; extracellular matrix; glycoprotein;
 F:1-23/Domain: signal sequence #status predicted <SIG>
 F:24-153/Domain: amino-terminal propeptide #status predicted <PRO>
 F:31-91/Domain: von Willebrand factor type C repeat homology <VMC>
 F:154-1221/Product: collagen alpha 1(III) chain #status predicted <MAT>
 F:154-167/Region: amino-terminal nonhelical telopeptide
 F:168-1196/Region: helical
 F:1091-1093/Region: cell attachment (R-G-D) motif
 F:1197-1221/Region: carboxyl-terminal nonhelical telopeptide
 F:1222-1466/Domain: carboxyl-terminal propeptide #status predicted <CPR>
 F:1238-1466/Domain: fibrillar collagen carboxyl-terminal homology <FCC>
 F:24/Modified site: pyroglutamate carboxylic acid (Gln) (in mature form) #status predic
 F:153-154/Cleavage site: Pro-Gln (procollagen N-endopeptidase) #status predicted
 F:154/Modified site: pyrrolidone carboxylic acid (Gln) (in mature form) #status predi
 F:161,11212/Modified site: allysine (Lys) #status predicted
 F:263,284,860,977,1106/Modified site: 5-hydroxylysine (Lys) #status experimental
 F:263/Binding site: carboxylate (Lys) (covalent) #status experimental
 F:584,1094/Modified site: 5-hydroxylysine (Lys) (partial) #status experimental
 F:948-949/Cleavage site: Gly-Ile (collagenase) #status experimental
 F:1106/Binding site: carboxylate (Lys) (covalent) #status predicted
 Query Match 65.2%; Score 60; DB 1; Length 1466;
 Best Local Similarity 78.6%; Pred. No. 0.57;
 Matches 11; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 1 GYCKGNPMPGA 14
 I I I I I I I I I I
 Db 738 GPKGDKGPGPGGA 751
 RESULT 23
 A61262
 collagen alpha 1(XVII) chain - human (fragment)
 N:Alternate names: bullous pemphigoid 180K autoantigen BPAG2; bullous pemphigoid anti
 C:Species: Homo sapiens (man)
 C:Date: 12-May-1994 #sequence, revision 12-Jul-1996 #text_change 21-Jul-2000
 C:Accession: I56325; I55345; A61262
 R:Giudice, G.J.; Emery, D.J.; Diaz, L.A.
 J. Invest. Dermatol. 99, 243-250, 1992
 A>Title: Cloning and primary structural analysis of the bullous pemphigoid autoantige
 A:Reference number: I56325; MUID:92381323
 A:Accession: I56325
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-1532 <RES>
 A:Cross-references: GB:M91669; NID:9179516; PIDN:AA35605.1; PID:9179517
 R:Li, K.H.; Sawamura, D.; Giudice, G.J.; Diaz, L.A.; Mattei, M.G.; Chu, M.L.; Utto,
 J. Biol. Chem. 266, 24064-24069, 1991
 A>Title: Genomic organization of collagenous domains and chromosomal assignment of hu
 A:Reference number: I55345; MUID:92084712
 A:Accession: I55345
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 543-890, 'P' <RE2>
 A:Cross-references: GB:M63730; NID:9179520; PIDN:AA51839.1; PID:9179521
 R:Giudice, G.J.; Squiquera, H.L.; Elias, P.M.; Diaz, L.A.
 J. Clin. Invest. 87, 734-738, 1991
 A>Title: Identification of two collagen domains within the bullous pemphigoid autoant
 A:Reference number: A61262; MUID:91123476
 A:Accession: A61262
 A:Molecule type: mRNA
 A:Residues: 543-890, 'P' <GIO>
 C:Genetics: GDB:COL17A1; BPAG2; BP180
 A:Gene: GDB:COL17A1; BPAG2; BP180

A;Cross-references: GDB:131396; OMIM:113811
A;Map position: 10q24.3-10q24.3

Query Match	65.28;	Score 60;	DB 2;	Length 1532;
Best Local Similarity	78.68;	Pred. No. 0.59;		
Matches 11; Conservative	1;	Mismatches 2;	Indels 0;	Gaps 0

```
QY      1 GVKGDKGNPGWPGA 14
          | | | | | : | | |
Db      995 GPKGDKGDPGVPGA 1008
```

RESULT 24
T13990
collagen type IV alpha 2 - fruit fly (*Drosophila melanogaster*)
C:Species: Drosophila melanogaster
C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 17-Nov-2000
C:Accession: T13990
R:Yaotianornsiikul, S.; Davis, W.J.; Cramer, G.; Kindrell, D.A.; Dearolf, C.R.
submitted to the EMBL Data Library, July 1996
A:Description: Viking: Identification and characterization of a novel type IV collagen
A:Reference number: Z17845
A:Accession: T13990
A:Status: preliminary; translated from GB/EMBL/DBD
A:Molecule type: mRNA
A:Residues: 1-1761 <YMS>
A:Cross-references: EMBL:U65431; NID:g2281290; PID:g2281291; PIDN:AAB64082.1
C:Genetics:
A:Gene: COL1A2
A:Cross-references: FlyBase:FBgn0016075
C:Superfamily: collagen alpha 1(IV) chain

Query Match	65.2%	Score 60;	DB 2;	Length 1761;
Best Local Similarity	66.7%	Pred. No. 0.68;		
Matches 10;	Conservative 2;	Mismatches 3;	Indels 0;	Gaps 0

```
Qy      1  GVKGDKNPGWPGAP  15
          | | : | | | : | |
Db      994 GRKGETGNPGFPGRP 1008
```

RESULT 25
A45407
collagen alpha 3(IV) chain - sea urchin (Strongylocentrotus purpuratus)
C:Species: Strongylocentrotus purpuratus (purple urchin)
C:Date: 22-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 13-Aug-1999
C:Accession: A45407; A43903; A23940
R:Exposito, J.Y.; D'Alessio, M.; Di Liberto, M.; Ramirez, F.
J. Biol. Chem. 268, 5249-5254, 1993
A:Title: Complete primary structure of a sea urchin type IV collagen alpha chain and an
A:Reference number: A45407; MUID:93186842
A:Accession: A45407
A:Status: preliminary; not compared with conceptual translation
A:Molecule type: nucleic acid
A:Residues: 1-1752 <EXP>
A:Note: sequence extracted from NCBI backbone (NCBIP:126841)
R:Wessel, G.M.; Etkin, M.; Benson, S.
Dev. Biol. 148, 261-272, 1991
A:Title: Primary mesenchyme cells of the sea urchin embryo require an autonomously prod
A:Reference number: A43903; MUID:92038439
A:Accession: A43903
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 'P', 633-1537, 'G', <MES>
A:Cross-references: G6:564572; NID:g238616; PID:AA82070.1; PID:g238617
A:Note: sequence extracted from NCBI backbone (NCBIN:64577, NCBIP:64573)
R:Venkatesan, M.; De Pablo, F.; Vogeli, G.; Simpson, R.T.
Proc. Natl. Acad. Sci. U.S.A. 83, 3351-3355, 1986
A:Title: Structure and developmentally regulated expression of a Strongylocentrotus purpur
A:Reference number: A23940; MUID:86205894
A:Accession: A23940

A:Molecule type: DNA
A:Residues: 742-812 <VEN>
A:Cross-references: EMBL:M13206
C:Superfamily: collagen alpha 1(IV) chain
C:Keywords: collod coll; extracellular matrix; glycoprotein; trimer; triple helix
F:29-161/Domain: amino-terminal nonhelical, 75 <7SD>
F:162-1523/Region: Interrupted helical
F:1524-1752/Domain: carboxyl-terminal nonhelical, NC1 <NC1>
F:1534-1634/Domain: collagen IV carboxyl-terminal repeat <CR1>
F:1644-1748/Domain: collagen IV carboxyl-terminal repeat <CR2>
F:129/Modified site: allysine (Lys) #status predicted

Query Match	64.7%	Score 59.5;	DB 2;	Length 1752;
Best Local Similarity	54.2%;	Pred. No. 0.8;		
Matches 13;	Conservative 0;	Mismatches 2;	Indels 9;	Gaps 1

QY 1 GVKGD-----KGNPGWEGAP 15
 ||||| ||||| |||
 Db 1177 GVKGDPRGTPEGAKGNPGLPGIP 1200

Search completed: May 23, 2001, 16:16:43
Job time: 84 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 23, 2001, 16:15:54 ; Search time 8.35 seconds

(without alignments)
61.537 Million cell updates/sec

Title: US-09-529-691-1

Perfect score: 92
Sequence: 1 GVKGDKGNPGWPGAP 15

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 93435 seqs, 34255486 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Database : SwisProt_39.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	92	100.0	1669	CA14_HUMAN	P02462 homo sapien
2	78	84.8	1669	CA14_MOUSE	P02463 mus musculu
3	69	75.0	1685	CA54_HUMAN	P29400 homo sapien
4	66	71.7	754	CA54_CANFA	P08122 mus musculu
5	65	70.7	1707	CA24_MOUSE	P08572 mus musculu
6	64	69.6	1712	CA24_HUMAN	P08572 mus musculu
7	63	68.5	2944	CA17_HUMAN	P02388 homo sapien
8	62	67.4	1670	CA34_HUMAN	P01955 homo sapien
9	61	66.3	339	COA5_BPRD	P22336 bacterioph
10	61	66.3	1603	CA1F_HUMAN	P07092 homo sapien
11	61	66.3	1690	CA44_HUMAN	P53420 homo sapien
12	61	66.3	1763	CA24_ASCSU	P27393 ascaris suu
13	60	65.2	369	PSPD_BOVIN	P35246 bos taurus
14	60	65.2	623	CA44_RABIT	P55787 oryctolagus
15	60	65.2	636	CA13_RAT	P13441 ratu
16	60	65.2	1049	CA13_BOVIN	P04258 bos taurus
17	60	65.2	1466	CA13_BOVIN	P02461 homo sapien
18	59	64.1	471	CA34_BOVIN	P02461 homo sapien
19	59	64.1	1758	CA14_CAEEL	P17139 caenorhabd
20	58	63.0	285	YRF2_CAEEL	P09578 caenorhabd
21	58	63.0	1516	CA1H_HUMAN	P39060 homo sapien
22	58	63.0	1758	CA24_CAEEL	P17140 caenorhabd
23	57	62.0	298	CC34_CAEEL	P34687 caenorhabd
24	56	60.9	1464	CA13_MOUSE	P08121 mus musculu
25	55.5	60.3	458	MSRE_MOUSE	P30204 mus musculu
26	55	59.8	245	CLOC_HUMAN	P02747 homo sapien
27	55	59.8	266	YMK_CAEEL	P21184 caenorhabd
28	55	59.8	371	COM6_BOVIN	P23305 bos taurus
29	55	59.8	680	CA1A_MOUSE	P00506 mus musculu
30	55	59.8	744	CA18_HUMAN	P27558 homo sapien
31	55	59.8	921	CA19_MOUSE	P05722 mus musculu
32	55	59.8	963	YQ36_CAEEL	P09457 caenorhabd
33	55	59.8	1650	CA2B_MOUSE	P064739 mus musculu

ALIGNMENTS

34	55	59.8	1678	1	CA64_HUMAN	Q14031 homo sapien
35	55	59.8	1736	1	CA2B_HUMAN	P13942 homo sapien
36	54	58.7	316	1	CC12_CAEEL	P20630 caenorhabd
37	54	58.7	316	1	CC13_CAEEL	P20631 caenorhabd
38	54	58.7	353	1	CA29_CHICK	P12108 gallus gall
39	54	58.7	375	1	PSPD_HUMAN	P35247 homo sapien
40	54	58.7	635	1	CA2B_HUMAN	P25067 homo sapien
41	54	58.7	674	1	CA1A_CHICK	P25067 homo sapien
42	54	58.7	675	1	CA39_CHICK	P32017 gallus gall
43	54	58.7	743	1	CA18_MOUSE	P00780 mus musculu
44	54	58.7	744	1	CA18_RABIT	P14282 oryctolagus
45	54	58.7	1775	1	CA14_DROME	P08120 drosophila

RESULT 1	
ID	CA14_HUMAN
AC	P02462
DT	21-JUL-1986 (Rel. 01, Created)
DT	01-FEB-1996 (Rel. 33, Last sequence update)
DT	01-OCT-2000 (Rel. 40, Last annotation update)
DE	COLLAGEN ALPHA 1(IV) CHAIN PRECURSOR.
GN	COL4A1.
OS	Homo sapiens (Human).
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX	NCBI_TaxID=9606;
RN	(1)
RP	SEQUENCE FROM N.A.
RX	MEDLINE=89340433; PubMed=2701944;
RA	Solomon R., Huotari M., Ganguly A., Prockop D.J., Trygvason K.;
RT	"Structural organization of the gene for the alpha 1 chain of human
RT	type IV collagen.";
RT	J. Biol. Chem. 264:13565-13571(1989).
RN	(2)
RP	SEQUENCE OF 46-1257 FROM N.A.
RC	TISSUE=Placenta;
RX	MEDLINE=88083584; PubMed=3691802;
RA	Solomon R., Haka-Risku T., Prockop D.J., Trygvason K.;
RT	"Complete primary structure of the alpha 1-chain of human basement
RT	membrane (type IV) collagen.";
RT	FEBS Lett. 225:188-194(1987).
RN	(3)
RP	SEQUENCE OF 1-943 FROM N.A.
RC	TISSUE=Placenta;
RX	MEDLINE=88029471; PubMed=3311751;
RA	Brazel D., Oberhauser I., Dieringer H., Babel W., Glanville R.W.,
RA	Deutmann R., Kuehn K.;
RT	"Completion of the amino acid sequence of the alpha 1 chain of human
RT	basement membrane collagen (type IV) reveals 21 non-triplet
RT	interruptions located within the collagenous domain.";
RT	Eur. J. Biochem. 168:529-536(1987).
RN	[4]
RP	SEQUENCE OF 28-243.
RX	MEDLINE=86004708; PubMed=4043082;
RA	Glanville R.W., Qian R.O., Stebold B., Ristell J., Kuehn K.;
RT	"Amino acid sequence of the N-terminal aggregation and cross-linking
RT	region (75 domain) of the alpha 1 (IV) chain of human basement
RT	membrane collagen.";
RT	Eur. J. Biochem. 152:213-219(1985).
RN	[5]
RP	SEQUENCE OF 534-1447.
RX	MEDLINE=85003629; PubMed=6434307;
RA	Babel W., Glanville R.W.;
RT	"Structure of human basement-membrane (type IV) collagen. Complete,
RT	amino-acid sequence of a 914-residue-long pepsin fragment from the
RT	alpha 1(IV) chain.";
RT	Eur. J. Biochem. 143:545-556(1984).
RN	[6]
RP	SEQUENCE OF 1256-1669 FROM N.A.

RX MEDLINE-85207819; PubMed-2581969;
 RA Pihlajaniemi T., Tryggvason K., Myers J.C., Kurkinen M., Lebo R.,
 RA Cheung M.-C., Prockop D.J., Boyd C.D.;
 RT "cDNA clones coding for the pro-alpha1(IV) chain of human type IV
 RT procollagen reveal an unusual homology of amino acid sequences in two
 RT halves of the carboxyl-terminal domain.";
 RL J. Biol. Chem. 260:7681-7687(1985).
 RN [7]
 RP SEQUENCE OF 1259-1669 FROM N.A.
 RX MEDLINE-85216555; PubMed-2582422;
 RA Brinker J.M., Gudes L.J., Loidl H.R., Wang S.-Y., Rosenbloom J.,
 RA Kefauver N.A., Myers J.C.;
 RT "Restricted homology between human alpha 1 type IV and other
 RT procollagen chains.";
 RL Proc. Natl. Acad. Sci. U.S.A. 82:3649-3653(1985).
 RN [8]
 RP SEQUENCE OF 1-28 FROM N.A.
 RX MEDLINE-89034231; PubMed-3182844;
 RA Solhunen R., Huotari M., Hostalka S.L., Prockop D.J., Tryggvason K.;
 RT "The structural genes for alpha 1 and alpha 2 chains of human type IV
 RT collagen are divergently encoded on opposite DNA strands and have an
 RT overlapping promoter region.";
 RL J. Biol. Chem. 263:17217-17220(1988).
 RN [9]
 RP SEQUENCE OF 1441-1669, AND DISULFIDE BONDS.
 RX TISSUE-Placenta;
 RC MEDLINE-89005112; PubMed-2844531;
 RA Siebold B., Deutzmann R., Kuehn K.;
 RT "The arrangement of intra- and intermolecular disulfide bonds in the
 RT carboxyterminal, non-collagenous aggregation and cross-linking domain
 RT of basement-membrane type IV collagen.";
 RL Eur. J. Biochem. 176:617-624(1988).
 CC -1- FUNCTION: TYPE IV COLLAGEN IS THE MAJOR STRUCTURAL COMPONENT OF
 CC GLOMERULAR BASEMENT MEMBRANES (GBM), FORMING A 'CHICKEN-WIRE'
 CC MESHWORK TOGETHER WITH LAMININS, PROTEOGLYCAN AND ENACTIN/
 CC NITROGEN.
 CC -1- SUBUNIT: THERE ARE SIX TYPE IV COLLAGEN ISOFORMS, ALPHA 1(IV)-
 CC ALPHA 6(IV), EACH OF WHICH CAN FORM A TRIPLE HELIX STRUCTURE
 CC WITH 2 OTHER CHAINS TO GENERATE TYPE IV COLLAGEN NETWORK.
 CC -1- DOMAIN: ALPHA CHAINS OF TYPE IV COLLAGEN HAVE A NONCOLLAGENOUS
 CC DOMAIN (NC1) AT THEIR C-TERMINUS, FREQUENT INTERRUPTIONS OF THE
 CC G-X-Y REPEATS IN THE LONG CENTRAL TRIPLE-HELICAL DOMAIN (WHICH MAY
 CC CAUSE FLEXIBILITY IN THE TRIPLE HELIX), AND A SHORT N-TERMINAL
 CC TRIPLE-HELICAL /S DOMAIN.
 CC -1- PTM: LYSINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
 CC UNIT ARE HYDROXYLATED IN ALL CASES AND BIND CARBOHYDRATE.
 CC -1- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
 CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
 CC -1- PTM: TYPE IV COLLAGENS CONTAIN NUMEROUS CYSTEINE RESIDUES WHICH
 CC ARE INVOLVED IN INTER- AND INTRAMOLECULAR DISULFIDE BONDING. 12 OF
 CC THESE, LOCATED IN THE NC1 DOMAIN, ARE CONSERVED IN ALL KNOWN TYPE
 CC IV COLLAGENS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC DR EMBL; M26548; AAA53098.1; JOINED.
 CC DR EMBL; M26549; AAA53098.1; JOINED.
 CC DR EMBL; M26551; AAA53098.1; JOINED.
 CC DR EMBL; M26552; AAA53098.1; JOINED.
 CC DR EMBL; M26553; AAA53098.1; JOINED.
 CC DR EMBL; M26554; AAA53098.1; JOINED.
 CC DR EMBL; M26555; AAA53098.1; JOINED.
 CC DR EMBL; M26556; AAA53098.1; JOINED.
 CC DR EMBL; M26557; AAA53098.1; JOINED.
 CC DR EMBL; M26559; AAA53098.1; JOINED.
 CC DR EMBL; M26560; AAA53098.1; JOINED.
 CC DR EMBL; M26561; AAA53098.1; JOINED.
 CC DR EMBL; M26562; AAA53098.1; JOINED.
 CC DR EMBL; M26563; AAA53098.1; JOINED.
 CC DR EMBL; M26564; AAA53098.1; JOINED.
 CC DR EMBL; M26565; AAA53098.1; JOINED.
 CC DR EMBL; M26566; AAA53098.1; JOINED.
 CC DR EMBL; M26567; AAA53098.1; JOINED.
 CC DR EMBL; M26568; AAA53098.1; JOINED.
 CC DR EMBL; M26569; AAA53098.1; JOINED.
 CC DR EMBL; M26570; AAA53098.1; JOINED.
 CC DR EMBL; M26571; AAA53098.1; JOINED.
 CC DR EMBL; M26572; AAA53098.1; JOINED.
 CC DR EMBL; M26573; AAA53098.1; JOINED.
 CC DR EMBL; M26574; AAA53098.1; JOINED.
 CC DR EMBL; M26575; AAA53098.1; JOINED.
 CC DR EMBL; Y00706; CAA68698.1; -.
 CC DR EMBL; X05561; CAA29075.1; -.
 CC DR EMBL; M10940; AAA52006.1; -.
 CC DR EMBL; M11315; AAA52042.1; -.
 CC DR PIR; S16876; CGH4B.
 CC DR MIM; 120130; -.
 CC DR InterPro; IPR000087; -.
 CC DR InterPro; IPR001442; -.
 CC DR Pfam; PF01413; C4; 2.
 CC DR Pfam; PF01391; Collagen; 21.
 CC KW Extracellular matrix; Connective tissue; Basement membrane;
 CC Repeat; Hydroxylation; Glycoprotein; Collagen; Signal.
 CC KW
 CC FT SIGNAL 1 27
 CC FT PROPEP 28 172
 CC FT CHAIN 173 1669
 CC FT DOMAIN 173 1440
 CC FT CAROHD 1441 1669
 CC FT CAROHD 126 126
 CC FT DISULFD 1460 1551
 CC FT DISULFD 1493 1548
 CC FT DISULFD 1505 1511
 CC FT DISULFD 1570 1665
 CC FT DISULFD 1604 1662
 CC FT DISULFD 1616 1622
 CC FT CONFLCT 237 238
 CC FT CONFLCT 241 241
 CC FT CONFLCT 319 319
 CC FT CONFLCT 719 719
 CC FT CONFLCT 837 837
 CC FT CONFLCT 842 842
 CC FT CONFLCT 896 896
 CC FT CONFLCT 904 904
 CC FT CONFLCT 914 914
 CC FT CONFLCT 998 998
 CC FT CONFLCT 1010 1010
 CC FT CONFLCT 1012 1012
 CC FT CONFLCT 1358 1358
 CC SQ SEQUENCE 1669 AA; 160611 MW; 3BBBA6DFFB9B8A84 CRC64;

Query Match 100.0%; Score 92; DB 1; Length 1669;
 Best Local Similarity 100.0%; Pred. No. 5; De-06;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;


```

FT CONFLICT 26 26 A -> P (IN REF. 2).
FT CONFLICT 186 186 S -> L (IN REF. 2).
FT CONFLICT 319 319 Q -> S (IN REF. 2).
FT CONFLICT 369 369 Q -> L (IN REF. 2).
FT CONFLICT 403 403 L -> F (IN REF. 2).
FT CONFLICT 481 481 P -> L (IN REF. 2).
FT CONFLICT 493 493 Q -> H (IN REF. 2).
FT CONFLICT 712 712 S -> I (IN REF. 2).
FT CONFLICT 813 813 E -> Q (IN REF. 2).
FT CONFLICT 982 982 Q -> H (IN REF. 2).
FT CONFLICT 1397 1397 V -> S (IN REF. 3).
SQ SEQUENCE 1669 AA; 160680 MM; 42916B91E52058E9 CRC64;

Query Match      84.8%; Score 78; DB 1; Length 1669;
Best Local Similarity 86.7%; Pred. No. 0.00066;
Matches 13; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GVGKDKGNPGWPGAP 15
   1 111111 11111
DB 1263 GPGKDKGNPGWPGAP 1277

RESULT 3
CA54_HUMAN STANDARD; PRT; 1685 AA.
ID CA54_HUMAN STANDARD; PRT; 1685 AA.
AC P29400; Q16126; Q16006;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE COLLAGEN ALPHA 5(IV) CHAIN PRECURSOR.
GN COL4A5.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
OX NCBI_Taxid:9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-94165049; PubMed-8120014;
RA Zhou J., Leinonen A., Tryggvason K.;
RT "Structure of the human type IV collagen COL4A5 gene.";
RL J. Biol. Chem. 269:6608-6614(1994).
RN [2]
RP SEQUENCE OF 1-910 FROM N.A., AND VARIANT AS CYS-521.
RC TISSUE-Kidney;
RX MEDLINE-92316923; PubMed-1352287;
RA Zhou J., Hertz J.M., Leinonen A., Tryggvason K.;
RT "Complete amino acid sequence of the human alpha 5 (IV) collagen chain and identification of a single-base mutation in exon 23 converting glycine 521 in the collagenous domain to cysteine in an Alport syndrome patient.";
RL J. Biol. Chem. 267:12475-12481(1992).
RN [3]
RP SEQUENCE OF 85-1685 FROM N.A.
RC TISSUE-Placenta;
RX MEDLINE-90337990; PubMed-2380186;
RA Pihlajaniemi T., Pihlajaniemi E.R., Myers J.C.;
RT "Complete primary structure of the triple-helical region and the carboxyl-terminal domain of a new type IV collagen chain, alpha 5(IV).";
RL J. Biol. Chem. 265:13758-13766(1990).
RN [4]
RP SEQUENCE OF 924-1685 FROM N.A.
RX MEDLINE-91169491; PubMed-2004755;
RA Zhou J., Hostikka S.L., Chow L.T., Tryggvason K.;
RT "Characterization of the 3' half of the human type IV collagen alpha 5 gene that is affected in the Alport syndrome.";
RL Genomics 9:1-9(1991).
RN [5]
RP SEQUENCE OF 914-1685 FROM N.A.
RX MEDLINE-90160375; PubMed-1689491;
RA Hostikka S.L., Eddy R.L., Byers M.G., Hoeyhtyae M., Shows T.B., Tryggvason K.;

```

```

RT "Identification of a distinct type IV collagen alpha chain with restricted kidney distribution and assignment of its gene to the locus of X chromosome-linked Alport syndrome.";
RT Proc. Natl. Acad. Sci. U.S.A. 87:1606-1610(1990).
RN [6]
RP SEQUENCE OF 1442-1471 FROM N.A.
RX MEDLINE-90252791; PubMed-2339699;
RA Myers J.C., Jones T.A., Pihlajaniemi E.R., Kadri A.S., Goddard A.D., Sheer D., Solomon E., Pihlajaniemi T.;
RT "Molecular cloning of alpha 5(IV) collagen and assignment of the gene to the region of the X chromosome containing the Alport syndrome locus.";
RL Am. J. Hum. Genet. 46:1024-1033(1990).
RN [7]
RP SEQUENCE OF 1-20 FROM N.A.
RA Guo C., van Damme B., Vanrenterghem Y., Devriendt K., Cassiman J., Marynen P.;
RT Submitted (SEP-1994) to the EMBL/GenBank/DBJ databases.
RN [8]
RP SEQUENCE OF 1258-1270 FROM N.A. (SPLICED FORM).
RX MEDLINE-94133540; PubMed-8301933;
RA Guo C., van Damme B., van Damme-Lombaerts R., van den Berghe H., Cassiman J.-J., Marynen P.;
RT "Differential splicing of COL4A5 mRNA in kidney and white blood cells: a complex mutation in the COL4A5 gene of an Alport patient deletes the NC1 domain.";
RL Kidney Int. 44:1316-1321(1993).
RN [9]
RP REVIEW ON VARIANTS.
RX MEDLINE-97338662; PubMed-9195222;
RA Lemmink H.H., Schroeder C.H., Monnens L.A.H., Smeets H.J.M.;
RT "The clinical spectrum of type IV collagen mutations.";
RL Hum. Mutat. 9:477-499(1997).
RN [10]
RP VARIANT AS SER-1564.
RX MEDLINE-91169492; PubMed-1672282;
RA Zhou J., Barker D.F., Hostikka S.L., Gregory M.C., Atkin C.L., Tryggvason K.;
RT "Single base mutation in alpha 5(IV) collagen chain gene converting a conserved cysteine to serine in Alport syndrome.";
RL Genomics 9:10-18(1991).
RN [11]
RP VARIANT AS ARG-325.
RX MEDLINE-92303559; PubMed-1376965;
RA Knebelmann B., Deschenes G., Gros F., Hors M.-C., Grunfeld J.-P., Tryggvason K., Gubler M.-C., Antignac C.;
RT "Substitution of arginine for glycine 325 in the collagen alpha 5 (IV) chain associated with X-linked Alport syndrome: characterization of the mutation by direct sequencing of PCR-amplified lymphoblast cDNA fragments.";
RL Am. J. Hum. Genet. 51:135-142(1992).
RN [12]
RP VARIANT AS GLU-325.
RX MEDLINE-93244772; PubMed-1363780;
RA Renieri A., Seri M., Myers J.C., Pihlajaniemi T., Massella L., Rizzoni G., de Marchi M.;
RT "De novo mutation in the COL4A5 gene converting glycine 325 to glutamic acid in Alport syndrome.";
RL Hum. Mol. Genet. 1:127-129(1992).
RN [13]
RP VARIANTS AS THR-1517, SER-1538 AND GLN-1563.
RX MEDLINE-94010948; PubMed-8406498;
RA Lemmink H.L., Schroeder C.H., Brunner H.G., Nelen M.R., Zhou J., Tryggvason K., Haggas-Schouten W.A.G., Roodvoets A.P., Rascher W., van Oost B.A., Smeets H.J.M.;
RT "Identification of four novel mutations in the COL4A5 gene of patients with Alport syndrome.";
RL Genomics 17:485-489(1993).
RN [14]
RP VARIANTS AS E-400/V-406/V-638/A-638/R-653/R-796/R-869/R-872 & C-1241.
RX MEDLINE-95322976; PubMed-7596631;
RA Boye E., Flinter F., Zhou J., Tryggvason K., Bobrow M., Harris A.;
RT "Detection of 12 novel mutations in the collagenous domain of the

```

RT COL4A5 gene in Alport syndrome patients.";
 RL Hum. Mutat. 5:197-204(1995).
 RN [15]
 RP VARIANT AS ARG-1649.
 RX MEDLINE-96213750; PubMed-8651292;
 RA Barker D.F., Pruchno C.J., Jiang X., Atkin C.L., Stone E.M.,
 RT "A mutation causing Alport syndrome with tardive hearing loss is
 common in the western United States.";
 RL Am. J. Hum. Genet. 58:1157-1165(1996).
 RN [16]
 RP VARIANTS AS
 RX MEDLINE-96213754; PubMed-8651296;
 RA Renieri A., Bruttini M., Galli L., Zanelli P., Neri T., Rossetti S.,
 RA Tuerker A., Heiskari N., Zhou J., Gusmano R., Maseella L., Banfi G.,
 RA Scolari F., Sessa A., Rizzoni G., Tryggvason K., Pignatelli P.F.,
 RA Savi M., Ballabio A., de Marchi M.;
 RT "X-linked Alport syndrome: an SSCP-based mutation survey over all 51
 exons of the COL4A5 gene.";
 RL Am. J. Hum. Genet. 58:1192-1204(1996).
 RN [17]
 RP VARIANT AS ASP-1498.
 RX MEDLINE-96233932; PubMed-8829632;
 RA Tverekaya S., Bobrylna V., Tsalykova F., Ignatova M.,
 RA Krasnopol'skaya X., Evgrafov O.;
 RT "Substitution of A1498D in noncollagen domain of $\alpha 5(\text{IV})$ collagen
 chain associated with adult-onset X-linked Alport syndrome.";
 RL Hum. Mutat. 7:149-150(1996).
 RN [18]
 RP VARIANT AS GLN-1677.
 RX MEDLINE-97295089; PubMed-9150741;
 RA Barker D.F., Denison J.C., Atkin C.L., Gregory M.C.;
 RT "Common ancestry of three Ashkenazi-American families with Alport
 syndrome and COL4A5 R1677Q.";
 RL Hum. Genet. 99:681-684(1997).
 RN [19]
 RP VARIANTS AS R-174; R-177; R-325; C-1410; W-1421; T-1517 AND D-1596.
 RX MEDLINE-98112435; PubMed-9452056;
 RA Neri T.M., Zanelli P., de Palma G., Savi M., Rossetti S., Turco A.E.,
 RA Pignatelli G.F., Galli L., Bruttini M., Renieri A., Mingarelli R.,
 RA Tivell A., Pinciaroli A.R., Ragazzo M., Rizzoni G.F., de Marchi M.;
 RT "Missense mutations in the COL4A5 gene in patients with X-linked
 Alport syndrome.";
 RL Hum. Mutat. Suppl. 1:S106-S109(1998).
 RN [20]
 RP VARIANTS AS, AND VARIANTS.
 RX MEDLINE-99140256; PubMed-10094548;
 RA Plant K.E., Green P.M., Vetric D., Flinter F.A.;
 RT "Detection of mutations in COL4A5 in patients with Alport syndrome.";
 RL Hum. Mutat. 13:124-132(1999).
 CC -1- FUNCTION: TYPE IV COLLAGEN IS THE MAJOR STRUCTURAL COMPONENT OF
 GLOMERULAR BASEMENT MEMBRANES (GBM), FORMING A 'CHICKEN-WIRE'
 MESHWORK TOGETHER WITH LAMININS, PROTEOGLYCANS AND ENACTIN/
 NIDOGEN.
 CC -1- SUBUNIT: THERE ARE SIX TYPE IV COLLAGEN ISOFORMS, ALPHA 1(IV)-
 ALPHA 6(IV), EACH OF WHICH CAN FORM A TRIPLE HELIX STRUCTURE
 WITH 2 OTHER CHAINS TO GENERATE TYPE IV COLLAGEN NETWORK.
 CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS: 1 (SHOWN HERE) AND 2; ARE
 PROBABLY PRODUCED BY ALTERNATIVE SPLICING. ISOFORM 2 IS A LONGER
 ISOFORM FOUND IN KIDNEY, IN WHICH 2 EXTRA G-X-X REPEATS INTO THE
 TRIPLE-HELIX DOMAIN ARE INTRODUCED.
 CC -1- DOMAIN: ALPHA CHAINS OF TYPE IV COLLAGEN HAVE A NONCOLLAGENOUS
 G-X-Y REPEATS IN THE LONG CENTRAL TRIPLE-HELICAL DOMAIN (WHICH MAY
 CAUSE FLEXIBILITY IN THE TRIPLE HELIX), AND A SHORT N-TERMINAL
 TRIPLE-HELICAL 7S DOMAIN.
 CC -1- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
 UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
 CC -1- PTM: TYPE IV COLLAGENS CONTAIN NUMEROUS CYSTEINE RESIDUES WHICH
 ARE INVOLVED IN INTER- AND INTRAMOLECULAR DISULFIDE BONDING. 12 OF
 THESE, LOCATED IN THE NCI DOMAIN, ARE CONSERVED IN ALL KNOWN TYPE
 IV COLLAGENS.
 CC -1- DISEASE: DEFECTS IN COL4A5 ARE ASSOCIATED WITH X-LINKED ALPORT

CC SYNDROME (AS). ALPORT SYNDROME IS CHARACTERIZED BY PROGRESSIVE
 CC GLOMERULONEPHRITIS, OFTEN ASSOCIATED WITH HIGH-TONE SENSORINEURAL
 CC DEAFNESS, SPECIFIC EYE ABNORMALITIES (LENTICONS AND MACULAR
 CC FLECKS), AND GLOMERULAR BASEMENT MEMBRANE DEFECTS. IN MALES, THE
 CC TYPICAL TIME COURSE FOR THE PROGRESS OF ALPORT SYNDROME IS:
 CC HEMATURIA BY THE AGE OF 5 YEARS, DEAFNESS AND HYPERTENSION IN
 CC
 Query Match 75.0%; Score 69; DB 1; Length 1685;
 Best local similarity 73.3%; Pred. No. 0.014;
 Matches 11; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
 QY 1 GYKDKNGNPGMPGAP 15
 DB 1279 GIKGKGNPGOPGUP 1293
 ID CAS4_CANFA STANDARD; PRT; 754 AA.
 AC Q28247;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-JUL-1998 (Rel. 36, Last annotation update)
 DE COLLAGEN ALPHA 5(IV) CHAIN (FRAGMENT).
 GN COL4A5.
 OS Canis familiaris (Dog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
 OX NCBI_TaxID=9615;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=SAMOYED; TISSUE=Kidney;
 RX MEDLINE-94224868; PubMed-8171024;
 RA Zheng K., Thorne P.S., Mariano P., Bauman R., McInnes R.R.;
 RT "Canine X chromosome-linked hereditary nephritis: a genetic model for
 human X-linked hereditary nephritis resulting from a single base
 mutation in the gene encoding the alpha 5 chain of collagen type
 IV.";
 RL Proc. Natl. Acad. Sci. U.S.A. 91:3989-3993(1994).
 CC -1- FUNCTION: TYPE IV COLLAGEN IS THE MAJOR STRUCTURAL COMPONENT OF
 GLOMERULAR BASEMENT MEMBRANES (GBM), FORMING A 'CHICKEN-WIRE'
 MESHWORK TOGETHER WITH LAMININS, PROTEOGLYCANS AND ENACTIN/
 NIDOGEN.
 CC -1- SUBUNIT: THERE ARE SIX TYPE IV COLLAGEN ISOFORMS, ALPHA 1(IV)-
 ALPHA 6(IV), EACH OF WHICH CAN FORM A TRIPLE HELIX STRUCTURE
 WITH 2 OTHER CHAINS TO GENERATE TYPE IV COLLAGEN NETWORK.
 CC -1- SUBCELLULAR LOCATION: CELL SURFACE (POTENTIAL).
 CC -1- DOMAIN: ALPHA CHAINS OF TYPE IV COLLAGEN HAVE A NONCOLLAGENOUS
 G-X-Y REPEATS IN THE LONG CENTRAL TRIPLE-HELICAL DOMAIN (WHICH MAY
 CAUSE FLEXIBILITY IN THE TRIPLE HELIX), AND A SHORT N-TERMINAL
 TRIPLE-HELICAL 7S DOMAIN.
 CC -1- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
 UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
 CC -1- PTM: TYPE IV COLLAGENS CONTAIN NUMEROUS CYSTEINE RESIDUES WHICH
 ARE INVOLVED IN INTER- AND INTRAMOLECULAR DISULFIDE BONDING. 12 OF
 THESE, LOCATED IN THE NCI DOMAIN, ARE CONSERVED IN ALL KNOWN TYPE
 IV COLLAGENS.
 CC -1- DISEASE: A DEFECT IN COL4A5 HAS BEEN FOUND TO BE THE CAUSE OF
 CANINE X-LINKED HEREDITARY NEPHRITIS (HN), A DISEASE SIMILAR TO
 THAT IN HUMANS (ALSO REFERRED TO AS ALPORT SYNDROME) CHARACTERIZED
 BY PROGRESSIVE RENAL FAILURE AND NEUROSENSORY DEAFNESS.
 CC -1- SIMILARITY: TO OTHER TYPE IV COLLAGENS.
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: U07888; AAB60258.1; -.

DR Pfam: PF01391; Collagen: 19.
 KW Extracellular matrix; Connective tissue; Repeat: Hydroxylation;
 KM Glycoprotein; Basement membrane; Collagen; Signal.
 FT SIGNAL 1 25
 FT PROPEP 26 183
 FT CHAIN 184 1707
 FT DOMAIN 184 1479
 FT DOMAIN 1480 1707
 FT DISULFID 1499 1588
 FT DISULFID 1532 1585
 FT DISULFID 1544 1550
 FT DISULFID 1607 1703
 FT DISULFID 1641 1700
 FT DISULFID 1653 1660
 FT CARBOHYD 138 138
 FT CARBOHYD 1270 1270
 FT CONFLICT 1051 1051
 FT CONFLICT 1097 1097
 FT CONFLICT 1171 1171
 FT CONFLICT 1179 1179
 FT CONFLICT 1241 1241
 FT CONFLICT 1328 1328
 FT CONFLICT 1573 1573
 FT CONFLICT 1623 1623
 FT SEQUENCE 1707 AA; 167391 MW; 1A565159605FD508 CRC64;
 SO
 Query Match 70.7%; Score 65; DB 1; Length 1707;
 Best Local Similarity 66.7%; Pred. No. 0.056;
 Matches 10; Conservative 2; Mismatches 3; Indels 0; Gaps 0;
 QY 1 GYAKDKGNPGMPCAP 15
 ID 1165 GIPGDKGDFGMPGPV 1179
 DB
 RESULT 6
 CA24_HUMAN STANDARD; PRT; 1712 AA.
 AC P08572;
 DT 01-AUG-1988 (Rel. 08, Created)
 DT 01-FEB-1991 (Rel. 17, Last sequence update)
 DT 01-OCT-2000 (Rel. 40, Last annotation update)
 DE COLLAGEN ALPHA 2(IV) CHAIN PRECURSOR.
 GN COL4A2.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA MEDLINE=8906769; PubMed=3198637;
 RX Hostlika S.L., Tryggvason K.;
 RT "The complete primary structure of the alpha 2 chain of human type IV
 collagen and comparison with the alpha 1(IV) chain.";
 RL J. Biol. Chem. 263:19488-19493(1988).
 RN [2]
 RP SEQUENCE OF 1-1042 FROM N.A.
 RC TISSUE-Placenta;
 RX MEDLINE=88151998; PubMed=3345760;
 RA Brazel D., Pollner R., Oberhauser I., Kuehn K.;
 RT "Human basement membrane collagen (type IV). The amino acid sequence
 of the alpha 2(IV) chain and its comparison with the alpha 1(IV)
 chain reveals deletions in the alpha 1(IV) chain.";
 RL Eur. J. Biochem. 172:35-42(1988).
 RN [3]
 RP SEQUENCE OF 1254-1712 FROM N.A.
 RX MEDLINE=87219158; PubMed=3582677;
 RA Hostlika S.L., Kurkinen M., Tryggvason K.;
 RT "Nucleotide sequence coding for the human type IV collagen alpha 2
 chain cDNA reveals extensive homology with the NC-1 domain of alpha 1
 (IV) but not with the collagenous domain or 3'-untranslated region.";
 RL FEBS Lett. 216:281-286(1987).

RN [4]
 RP SEQUENCE OF 1451-1485 FROM N.A.
 RX MEDLINE=87092438; PubMed=3025878;
 RA Griffin C.A., Emanuel B.S., Hansen J.R., Cavenee W.K., Myers J.C.;
 RT "Human collagen genes encoding basement membrane alpha 1 (IV) and
 alpha 2 (IV) chains map to the distal long arm of chromosome 13.";
 RL Proc. Natl. Acad. Sci. U.S.A. 84:512-516(1987).
 RN [5]
 RP SEQUENCE OF 1486-1712 FROM N.A.
 RX MEDLINE=87250571; PubMed=2439508;
 RA Myers J.C., Howard P.S., Jelen A.M., Dion A.S., Macarak E.J.;
 RT "Duplication of type IV collagen COOH-terminal repeats and species-
 specific expression of alpha 1(IV) and alpha 2(IV) collagen genes.";
 RL J. Biol. Chem. 262:9231-9238(1987).
 RN [6]
 RP SEQUENCE OF 1-33 FROM N.A.
 RX MEDLINE=89034231; PubMed=3182844;
 RA Solinen R., Huotari M., Hostlika S.L., Prockop D.J., Tryggvason K.;
 RT "The structural genes for alpha 1 and alpha 2 chains of human type IV
 collagen are divergently encoded on opposite DNA strands and have an
 overlapping promoter region.";
 RL J. Biol. Chem. 263:17217-17220(1988).
 RN [7]
 RP SEQUENCE OF 1-33 FROM N.A.
 RX MEDLINE=89030632; PubMed=2846280;
 RA Poeschl E., Pollner R., Kuehn K.;
 RT "The genes for the alpha 1(IV) and alpha 2(IV) chains of human
 basement membrane collagen type IV are arranged head-to-head and
 separated by a bidirectional promoter of unique structure.";
 RL EMBO J. 7:2687-2695(1988).
 RN [8]
 RP SEQUENCE OF 1-33 FROM N.A.
 RX TISSUE-Skin;
 RC MEDLINE=93305049; PubMed=8317999;
 RA Fischer G., Schmidt C., Opitz J., Cully Z., Kuehn K., Poeschl E.;
 RT "Identification of a novel sequence element in the common promoter
 region of human collagen type IV genes. Involved in the regulation of
 divergent transcription.";
 RL Biochem. J. 292:687-695(1993).
 RN [9]
 RP SEQUENCE OF 1480-1535; 1545-1614; 1617-1701 AND 1705-1712.
 RC TISSUE-Placenta;
 RX MEDLINE=89005112; PubMed=2844531;
 RA Siebold B., Deutzmann R., Kuehn K.;
 RT "The arrangement of intra- and intermolecular disulfide bonds in the
 carboxyterminal, non-collagenous aggregation and cross-linking domain
 of basement membrane type IV collagen.";
 RL Eur. J. Biochem. 176:617-624(1988).
 RN [10]
 RP FUNCTION: TYPE IV COLLAGEN IS THE MAJOR STRUCTURAL COMPONENT OF
 GLOMERULAR BASEMENT MEMBRANES (GBM), FORMING A 'CHICKEN-WIRE'
 MESHWORK TOGETHER WITH LAMININS, PROTEGLYCANS AND ENTRACTIN/
 NIIDGEN.
 CC
 CC -1- SUBUNIT: THERE ARE SIX TYPE IV COLLAGEN ISOFORMS, ALPHA 1(IV) -
 ALPHA 6(IV), EACH OF WHICH CAN FORM A TRIPLE HELIX STRUCTURE
 WITH 2 OTHER CHAINS TO GENERATE TYPE IV COLLAGEN NETWORK.
 CC -1- DOMAIN: ALPHA CHAINS OF TYPE IV COLLAGEN HAVE A NONCOLLAGENOUS
 CC DOMAIN (NC1) AT THEIR C-TERMINUS, FREQUENT INTERRUPTIONS OF THE
 CC G-X-Y REPEATS IN THE LONG CENTRAL TRIPLE-HELICAL DOMAIN (WHICH MAY
 CC CAUSE FLEXIBILITY IN THE TRIPLE HELIX), AND A SHORT N-TERMINAL
 CC TRIPLE-HELICAL 7S DOMAIN.
 CC
 CC -1- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPLEPEPTIDE REPEATING
 CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
 CC -1- PTM: TYPE IV COLLAGENS CONTAIN NUMEROUS CYSTEINE RESIDUES WHICH
 CC ARE INVOLVED IN INTER- AND INTRAMOLECULAR DISULFIDE BONDING. 12 OF
 CC THESE, LOCATED IN THE NC1 DOMAIN, ARE CONSERVED IN ALL KNOWN TYPE
 CC IV COLLAGENS.
 CC
 CC -----
 CC THIS SWISS-PROT entry is copyrighted. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)

```

CC or send an email to license@sib.ch).
CC EMBL: X05562; CAA29076.1; -
DR EMBL: X05610; CAA29098.1; -
DR EMBL: J02760; AAA58432.1; -
DR EMBL: M36963; AAA53099.1; -
DR EMBL: X12784; CAA31275.1; -
DR EMBL: J04217; AAA53097.1; -
DR PIR: A32024; A32024.
DR MIM: 120090; -
DR InterPro: IPR000087; -
DR InterPro: IPR001442; -
DR Pfam: PF01413; C4; 2.
DR Pfam: PF01391; Collagen; 20.
KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
KW Glycoprotein; Basement membrane; Collagen; Signal.
FT SIGNAL 1 25
FT PROPEP 26 183 AMINO-TERMINAL PROPEPTIDE (75 DOMAIN).
FT CHAIN 184 1712 COLLAGEN ALPHA 2 (IV) CHAIN.
FT DOMAIN 184 1484 TRIPLE-HELICAL REGION.
FT DOMAIN 1485 1712 NONHELICAL REGION (NC1).
FT DISULFID 1504 1593 OR 1590 (BY SIMILARITY).
FT DISULFID 1537 1590 OR 1593 (BY SIMILARITY).
FT DISULFID 1549 1555 BY SIMILARITY.
FT DISULFID 1612 1708 OR 1705 (BY SIMILARITY).
FT DISULFID 1646 1705 OR 1708 (BY SIMILARITY).
FT DISULFID 1658 1665 BY SIMILARITY.
FT CARBOHYD 138 138 N-LINKED (GLCNAC. . .).
FT CONFLICT 471 471 R -> P (IN REF. 2).
FT CONFLICT 683 683 A -> G (IN REF. 2).
FT CONFLICT 1575 1575 M -> I (IN REF. 5).
FT CONFLICT 1663 1663 G -> H (IN REF. 9).
FT CONFLICT 1701 1701 H -> G (IN REF. 9).
SQ SEQUENCE 1712 AA; 167535 MW; 2582A17847890037 CRC64;

Query Match 69.6%; Score 64; DB 1; Length 1712;
Best Local Similarity 73.3%; Pred. No. 0.079;
Matches 11; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 GYKGDGNGPMPGAP 15
Db 1368 GPKGPKDPPGAP 1382

RESULT 7
CA17_HUMAN
ID CA17_HUMAN STANDARD; PRT: 2944 AA.
AC 002388; Q14054; Q16507;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE COLLAGEN ALPHA 1(VII) CHAIN PRECURSOR (LONG-CHAIN COLLAGEN) (LC
CN COL7A1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94327588; PubMed=8051117;
RA Christianso A.M., Greenspan D.S., Lee S., Uitto J.;
RT "Cloning of human type VII collagen. Complete primary sequence of the
RT alpha 1(VII) chain and identification of intragenic polymorphisms.";
RL J Biol. Chem. 269:20256-20262(1994).
[2]
RP SEQUENCE OF 128-1493 FROM N.A., AND PARTIAL SEQUENCE.
RX MEDLINE=93338437; PubMed=1307247;
RA Christianso A.M., Rosenbaum L.M., Chung-Honet L.C., Parente M.G.,
RA Woodley D.T., Pan T.C., Zhang R.Z., Chu M.L., Burgeson R.E.,
RA Uitto J.;
RT "The large non-collagenous domain (NC-1) of type VII collagen is

```

RT Identification of a glycine-to-cysteine substitution in the triple-
 RL helical domain of type VII collagen."
 RL Hum. Mol. Genet. 4:1579-1583(1995).
 RN [12]
 RP VARIANT DDB ARG-2043.
 RX MEDLINE-95164985; PubMed-7861014;
 RA Christiano A.M., Moricone A., Paradisi M., Angelo C., Mazzanti C.,
 Cavalieri R., Uitto J.;
 RT "A glycine-to-arginine substitution in the triple-helical domain of
 RT type VII collagen in a family with dominant dystrophic epidermolysis
 RT bullosa."
 RL J. Invest. Dermatol. 104:438-440(1995).
 RN [13]
 RP VARIANTS RDEB AND DDB.
 RX MEDLINE-96220218; PubMed-8644729;
 RA Christiano A.M., McGrath J.A., Tan K.C., Uitto J.;
 RT "Glycine substitutions in the triple-helical region of type VII
 RT collagen result in a spectrum of dystrophic epidermolysis bullosa
 RT phenotypes and patterns of inheritance."
 RL Am. J. Hum. Genet. 58:671-681(1996).
 RN [14]
 RP VARIANT RDEB ARG-2575.
 RX MEDLINE-96154068; PubMed-8592061;
 RA Shmizu H., McGrath J.A., Christiano A.M., Nishikawa T., Uitto J.;
 RT "Molecular basis of recessive dystrophic epidermolysis bullosa:
 RT genotype/phenotype correlation in a case of moderate clinical
 RT severity."
 RL J. Invest. Dermatol. 106:119-124(1996).
 RN [15]
 RP VARIANT RDEB ARG-1782.
 RX MEDLINE-96183562; PubMed-8618018;
 RA Christiano A.M., McGrath J.A., Uitto J.;
 RT "Influence of the second COL7A1 mutation in determining the
 RT phenotypic severity of recessive dystrophic epidermolysis bullosa."
 RL J. Invest. Dermatol. 106:766-770(1996).
 RN [16]
 RP VARIANT RDEB ASP-2073.
 RX MEDLINE-96310789; PubMed-8757758;
 RA Dunnill M.G.S., McGrath J.A., Richards A.J., Christiano A.M.,
 Uitto J., Pope F.M., Eady R.A.J.;
 RT "Clinicopathological correlations of compound heterozygous COL7A1
 RT mutations in recessive dystrophic epidermolysis bullosa."
 RL J. Invest. Dermatol. 107:171-177(1996).
 RN [17]
 RP VARIANTS RDEB W-1982; G-2008; A-2025; E-2049; G-2063; W-2063 & R-2575.
 RX MEDLINE-97465605; PubMed-9326325;
 RA Hovenian A., Rochat A., Bodemer C., Petit E., Rivers C.A., Prost C.,
 Freitag S., Christiano A.M., Uitto J., Lathrop M., Barrandon Y.,
 de Prost Y.;
 RT "Characterization of 18 new mutations in COL7A1 in recessive
 RT dystrophic epidermolysis bullosa provides evidence for distinct
 RT molecular mechanisms underlying defective anchoring fibril
 RT formation."
 RL Am. J. Hum. Genet. 61:599-610(1997).
 CC -1- FUNCTION: STRATIFIED SQUAMOUS EPITHELIAL BASEMENT MEMBRANE PROTEIN
 CC THAT FORM ANCHORING FIBRILS WHICH MAY CONTRIBUTE TO EPITHELIAL
 CC BASEMENT MEMBRANE ORGANIZATION AND ADHERENCE BY INTERACTING WITH
 CC EXTRACELLULAR MATRIX (ECM) PROTEINS SUCH AS TYPE IV COLLAGEN.
 CC -1- SUBUNIT: HOMOTRIMER.
 CC -1- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
 CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
 CC -1- DISEASE: DEFECT IN COL7A1 IS A CAUSE OF RECESSIVE DYSTROPHIC
 CC EPIDERMOLYSIS BULLOSA (EB); A GROUP OF BLISTERING SKIN DISEASE
 CC CHARACTERIZED BY TISSUE SEPARATION WITHIN THE DERMAL-EPIDERMAL
 CC BASEMENT MEMBRANE ZONE. THE DYSTROPHIC TYPE, THE SCARRING FORM
 CC MANIFESTS THE CLEAVAGE WITHIN THE PAPILLARY DERMIS BELOW THE BASAL
 CC LAMINA AT THE LEVEL OF THE ANCHORING FIBRILS. BOTH AUTOSOMAL
 CC DOMINANT AND RECESSIVE INHERITANCE PATTERNS HAVE BEEN RECOGNIZED.
 CC -1- DISEASE: EPIDERMOLYSIS BULLOSA ACQUISITA (EBA) IS AN AUTOIMMUNE
 CC ACQUIRED BLISTERING SKIN DISEASE RESULTING FROM AUTOANTIBODIES TO
 CC TYPE VII COLLAGEN.
 CC -1- DISEASE: DEFECT IN COL7A1 IS A CAUSE OF DYSTROPHIC EPIDERMOLYSIS
 CC BULLOSA (DEB). DEB IS A BLISTERING SKIN DISEASES WITH EITHER AN

CC AUTOSOMAL DOMINANT (DDB) OR AUTOSOMAL RECESSIVE (RDEB)
 CC INHERITANCE. ULTRASTRUCTURALLY, DEB IS CHARACTERIZED BY TISSUE
 CC SEPARATION WHICH OCCURS BELOW THE DERMAL-EPIDERMAL BASEMENT
 CC MEMBRANE AT THE LEVEL OF THE ANCHORING FIBRILS. THE MOST SEVERE
 CC FORM OF RDEB IS THE HALLOPEAU-STEMENS TYPE (HS-RDEB), A Milder
 CC FORM IS THE MITIS TYPE (M-RDEB), AND THERE ARE STILL THE INVERSE,
 CC AND THE LOCALIZED TYPES. CLINICALLY, HS-RDEB MANIFESTS WITH
 CC MUTILATING SCARRING, WITH JOINT CONTRACTURES, CORNEAL EROSIONS,
 CC ESOPHAGUS STRUCTURES, AND PROPENSITY TO FORMATION OF CUTANEOUS
 CC SQUAMOUS CELL CARCINOMAS LEADING TO PREMATURE DEATH OF THE
 CC AFFECTED INDIVIDUALS. M-RDEB SHOWS LIFELONG BLISTERING TENDENCY,
 CC WITH LIMITED SCARRING AND LESS FREQUENT EXTRACUTANEOUS
 CC MANIFESTATIONS. HOWEVER THERE IS A CONTINUUM OF SEVERITY BETWEEN
 CC HS AND M-RDEB. THERE ARE SEVERAL TYPES OF DDB: THE COCKAYNE-
 CC TOURNAINE TYPE (CT-DDB), THE PASINI TYPE (P-DDB) WHICH IS SEVERE,
 CC BARTH TYPE (B-DDB) AND PRETIBIAL EPIDERMOLYSIS BULLOSA
 CC (PEB-DDB).
 CC -1- SIMILARITY: THE NC-1 DOMAIN OF TYPE VII COLLAGEN HAS HOMOLOGY
 CC TO FIBRONECTIN TYPE III DOMAINS, VON WILLEBRAND FACTOR A
 CC DOMAINS AND CARTILAGE MATRIX PROTEIN.
 CC -1- SIMILARITY: CONTAINS 1 PROTEASE INHIBITOR DOMAIN BELONGING TO THE
 CC BPTI/KUNITZ FAMILY OF INHIBITORS.
 CC -1- SIMILARITY: CONTAINS 9 FIBRONECTIN TYPE III-LIKE DOMAINS.
 CC -1- CAUTION: REF.5 SEQUENCE DIFFERS FROM THAT SHOWN IN POSITIONS
 CC 476 TO 523 DUE TO A FRAMESHIFT.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -

Query Match 68.5% Score 63; DB 1; Length 2944;
 Best Local Similarity 80.0%; Pred. No. 0.19;
 Matches 12; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 Oy 1 GYKGDKNPGWPGAP 15
 ||||| 111111
 Db 2386 GYKGDLPGLPGAP 2400
 RESULT 8
 CA34 HUMAN STANDARD; PRT; 1670 AA.
 ID CA34 HUMAN
 AC Q01955;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DT 01-OCT-2000 (Rel. 40, Last annotation update)
 DE COLLAGEN ALPHA 3(IV) CHAIN PRECURSOR.
 GN COL4A3.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_Taxid=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-Kidney;
 RX MEDLINE-94364994; PubMed-8083201;
 RA Maruyama M., Leinonen A., Mochizuki T., Tryggvason K., Reeder S.T.;
 RT "Complete primary structure of the human alpha 3(IV) collagen chain.
 RT Coexpression of the alpha 3(IV) and alpha 4(IV) collagen chains in
 RT human tissues."
 RL J. Biol. Chem. 269:23013-23017(1994).
 RN [2]
 RP REVISIONS:
 RA Leinonen A.;
 RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE OF 1386-1670 FROM N.A. AND PARTIAL SEQUENCE.
 RX MEDLINE-93015826; PubMed-1400291;
 RA Quiñones S., Bernat D., Garcia-Sogo M., Elena S.F., Sans J.;
 RT "Exon/intron structure of the human alpha 3(IV) gene encompassing the
 RT glycoprotein antigen (alpha 3(IV)NC1). Identification of a potentially
 RT antigenic region at the triple helix/NC1 domain junction."
 RL J. Biol. Chem. 267:19780-19784(1992).
 RN [4]

RP SEQUENCE OF 1453-1670 FROM N.A.
RX MEDLINE=91353570; PubMed=1802840;
RA Morrison K.E., Mariyama M., Yang-Feng T.L., Reeders S.T.;
RT "Sequence and localization of a partial cDNA encoding the human alpha
RT 3 chain of type IV collagen.";
RL Am. J. Hum. Genet. 49:545-554(1991).
RN [5]
RN SEQUENCE OF 1331-1670 FROM N.A.
RC TISSUE=Kidney;
RX MEDLINE=92147878; PubMed=1737849;
RA Turner N., Mason P.J., Brown R., Fox M., Povey S., Rees A.,
RA Pusey C.D.;
RT "Molecular cloning of the human Goodpasture antigen demonstrates it
RT to be the alpha 3 chain of type IV collagen.";
RL J. Clin. Invest. 89:592-601(1992).
RN [6]
RN SEQUENCE OF 1644-1670 FROM N.A.
RC TISSUE=Kidney;
RA Ding J.;
RT Submitted (JAN-1993) to the EMBL/Genbank/DBJ databases.
RN [7]
RN SEQUENCE OF 1439-1670, AND ALTERNATIVE SPLICING.
RC TISSUE=Kidney;
RX MEDLINE=94124597; PubMed=8294492;
RA Feng L., Xia Y., Wilson C.B.;
RT "Alternative splicing of the NC1 domain of the human alpha 3(IV)
RT collagen gene. Differential expression of mRNA transcripts that
RT predict three protein variants with distinct carboxyl regions.";
RL J. Biol. Chem. 269:2342-2348(1994).
RN [8]
RN SEQUENCE OF 1-29 FROM N.A.
RX MEDLINE=98196854; PubMed=9537506;
RA Momiota R., Sugimoto M., Oohashi T., Kigasawa K., Yoshioka H.,
RA Niimura Y.;
RT "Two genes, COL4A3 and COL4A4 coding for the human alpha3(IV) and
RT alpha4(IV) collagen chains are arranged head-to-head on chromosome
RL 2q36.";
RL FEBS Lett. 424:11-16(1998).
RN [9]
RN ALTERNATIVE SPLICING.
RX MEDLINE=93280184; PubMed=8505332;
RA Bernal D., Guinones S., Saus J.;
RT "The human mRNA encoding the Goodpasture antigen is alternatively
RT spliced.";
RL J. Biol. Chem. 268:12090-12094(1993).
RN [10]
RN VARIANT PRO-1474.
RX MEDLINE=95078827; PubMed=7967301;
RA Lemmink H.H., Mochizuki T., van den Heuvel L.P.W.J., Schroeder C.H.,
RA Barrientos A., Monnens L.A.H., van Oost B.A., Brunner H.G.,
RA Reeders S.T., Smeets H.J.M.;
RT "Mutations in the type IV collagen alpha 3 (COL4A3) gene in autosomal
RT recessive Alport syndrome.";
RL Hum. Mol. Genet. 3:1269-1273(1994).
CC -1 FUNCTION: TYPE IV COLLAGEN IS THE MAJOR STRUCTURAL COMPONENT OF
CC GLOMERULAR BASEMENT MEMBRANES (GBM), FORMING A 'CHICKEN-WIRE'
CC MESHWORK TOGETHER WITH LAMININS, PROTEOLYCAN AND ENACTIN/
CC NIDDOEN.
CC -1 SUBUNIT: THERE ARE SIX TYPE IV COLLAGEN ISOFORMS, ALPHA 1(IV)-
CC ALPHA 6(IV), EACH OF WHICH CAN FORM A TRIPLE HELIX STRUCTURE
CC WITH 2 OTHER CHAINS TO GENERATE TYPE IV COLLAGEN NETWORK.
CC -1 SUBCELLULAR LOCATION: CELL SURFACE (POTENTIAL).
CC -1 ALTERNATIVE PRODUCTS: AT LEAST 3 ISOFORMS; 1 (SHOWN HERE), 2/3 ARE
CC 3/4/5 ARE PRODUCED BY ALTERNATIVE SPLICING. THEY DIFFER IN THEIR
CC C-TERMINAL NC1 DOMAINS.
CC -1 TISSUE SPECIFICITY: ALPHA 3 AND ALPHA 4 TYPE IV COLLAGENS ARE
CC COLOCALIZED AND PRESENT ONLY IN BASEMENT MEMBRANES OF KIDNEY, EYE,
CC COCHLEA, LUNG AND BRAIN.
CC -1 DOMAIN: ALPHA CHAINS OF TYPE IV COLLAGEN HAVE A NONCOLLAGENOUS
CC DOMAIN (NC1) AT THEIR C-TERMINUS, FREQUENT INTERRUPTIONS OF THE
CC G-X-Y REPEATS IN THE LONG CENTRAL TRIPLE-HELICAL DOMAIN (WHICH MAY
CC CAUSE FLEXIBILITY IN THE TRIPLE HELIX), AND A SHORT N-TERMINAL
CC TRIPLE-HELICAL 7S DOMAIN.

CC	-1-	PM: PROLINS AT THE THIRD POSITION OF THE TRIPED REPEATING UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
CC	-1-	PM: THE ALTERNATIVE SPLICED FORM V CONTAINS AN ADDITIONAL N-LINKED GLYCOSYLATION SITE.
CC	-1-	PM: TYPE IV COLLAGENS CONTAIN NUMEROUS CYSTEINE RESIDUES WHICH ARE INVOLVED IN INTER- AND INTRA-MOLECULAR DISULFIDE BONDING. 12 OF THESE, LOCATED IN THE NC1 DOMAIN, ARE CONSERVED IN ALL KNOWN TYPE IV COLLAGENS.
CC	-1-	DISEASE: ANTIBODIES AGAINST THE NC1 DOMAIN OF ALPHA3(IV) MEDIATE THE HUMAN AUTOIMMUNE DISEASE, GOODPASTURE SYNDROME, WHICH IS CHARACTERIZED BY HEMATURIA AND PULMONARY HEMORRHAGE.
CC	-1-	DISEASE: DEFECTS IN COL4A3 ARE ASSOCIATED WITH THE TYPE I AUTOSOMAL RECESSIVE FORM OF ALPORT SYNDROME. AN HEREDITARY GLOMERULONEPHROPATHY CHARACTERIZED BY PROGRESSIVE RENAL FAILURE, HEMATURIA AND DEAFNESS. THE RECESSIVE FORM OCCURS EQUALLY BETWEEN MALES AND FEMALES.
CC	-1-	THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see http://www.isb.ch/announce/ or send an email to license@isb-sib.ch).
CC	EMBL:	X80031; CAA56335.1; -
DR	EMBL:	M92993; AAA21610.1; -
DR	EMBL:	S55790; AAB19637.1; -
DR	EMBL:	M81379; AAB51556.1; -
DR	EMBL:	L08650; AAB52044.1; -
DR	EMBL:	U02519; AAA18942.1; -
DR	EMBL:	U02520; AAA18943.1; -
DR	EMBL:	AB008495; BAA35064.1; -
DR	MIM:	120070; -
DR	MIM:	203780; -
DR	MIM:	233450; -
DR	InterPro:	IPR000087; -
DR	InterPro:	IPR001442; -
DR	Pfam:	PF01413; C4; 2.
DR	Pfam:	PF01391; Collagen; 21.
KW	Extracellular matrix:	Connective tissue; Repeat; Hydroxylation; Glycoprotein; Basement membrane; Collagen; Signal; Cell adhesion;
KM	Alternative splicing:	Polymorphism; Phosphorylation; POTENTIAL.
FT	SIGNAL	1 28
FT	CHAIN	29 1670
FT	DOMAIN	29 42
FT	DOMAIN	43 1438
FT	DOMAIN	1439 1670
FT	DOMAIN	1427 1444
FT	SITE	1426 1427
FT	SITE	791 793
FT	SITE	966 998
FT	SITE	1154 1156
FT	SITE	1306 1308
FT	SITE	1345 1347
FT	SITE	1432 1434
FT	SITE	253 253
FT	CARBOHYD	1435 1435
FT	MOD_RES	1437 1437
FT	MOD_RES	1460 1531
FT	DISULFID	1493 1548
FT	DISULFID	1505 1511
FT	DISULFID	1570 1665
FT	DISULFID	1604 1662
FT	DISULFID	1616 1622
FT	VARSPPLIC	1586 1670
FT	VARSPPLIC	1488 1670


```

FT FT MPNNAPITGRALEPIISRCTVCEGPALIAVHSSQTTDIP
FT CPGWISLWKGSFIMFTSAGSEGTQALASPSCSEFRA
FT SPFLCHGRGTCNYSNSYSPFVLASINPRMFKPIPSYVK
FT AGLEKTIISRCQVCMKRRH -> DALEFVVLNSP (IN
FT ISOFORM 3).
FT L-> P.
FT FT 1474 1474 /FTid-VAR_001908.
FT FT 1495 1495 O-> R.
FT FT VARIANT 1495 1495 /FTid-VAR_001909.
FT FT CONFLICT 1539 1539 R-> I (IN REF. 7).
FT FT CONFLICT 1594 1594 T-> A (IN REF. 4).
FT FT CONFLICT 1663 1664 QV-> HL (IN AAA52044 AND REF. 6).
FT FT SEQUENCE 1670 AA: 161740 MW: 9895708AF399BEAE CRC64:

Query Match 67.4% Score 62; DB 1; Length 1670;
Best Local Similarity 66.7% Pred. No. 0.15;
Matches 10; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 GYKDGKGNPGMGAP 15
Db 1189 GAKDGRGAPGPGPLP 1203

RESULT 9
ID COA5_BPPRD STANDARD; PRT: 339 AA.
AC P22536;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DT 01-DEC-1992 (Rel. 24, Last annotation update)
DE MINOR CAPSID PROTEIN (PROTEIN P5).
GN V.
OS Bacteriophage PRD1.
OC Viruses; dsDNA viruses, no RNA stage; Tectiviridae; Tectivirus.
OX NCBI_TaxID=10658;
RA MEDLINE-90320115; PubMed-2196741;
RA Bamford J.K.H., Bamford D.H.;
RT Capsomer proteins of bacteriophage PRD1, a bacterial virus with a
RT membrane.
RT Virology 177:445-451(1990).
RL [2]
RN COLLAGENOUS REGION.
RP MEDLINE-90206069; PubMed-2320123;
RA Bamford D.H., Bamford J.K.H.;
RT "Collagenous proteins multiply.";
RL Nature 344:497-497(1990).
CC -1- FUNCTION: THE SHORT COLLAGEN-LIKE REGION IS PROPOSED TO ACT AS A
CC TRIMERIZATION SIGNAL LEADING TO THE FULFILLMENT OF THE SYMMETRY
CC REQUIREMENT OF THE MINOR CAPSOMER.
CC -1- SUBUNIT: THE PRD1 VIRION IS COMPOSED OF AN ICOSAHERAL PROTEIN
CC COAT, AN INNER PROTEIN-LIPID MEMBRANE, AND A DSNA GENOME WHICH
CC IS LOCATED INSIDE THE LIPID VESICLE. THE PROTEIN COAT CONSISTS
CC MAINLY OF PROTEIN P3. THE MINOR PROTEIN P5 IS ALSO CONSIDERED TO
CC CONSTITUTE PART OF THE PROTEIN COAT.
CC -1- SIMILARITY: CONTAINS A COLLAGENOUS DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M69077; AAA32460.1; -
CC DR EMBL: M55568; AAA32446.1; -
CC DR PIR: B46345; B46345.
CC Coat protein.
FT INT_MET 0 0
FT DOMAIN 1 121 DOMAIN 1.

```

```

FT FT DOMAIN 122 140 COLLAGEN-LIKE.
FT FT DOMAIN 141 339 DOMAIN 2.
FT FT SEQUENCE 339 AA: 34318 MW: 184BCET7D08944F3 CRC64:

Query Match 66.3% Score 61; DB 1; Length 339;
Best Local Similarity 76.9% Pred. No. 0.046;
Matches 10; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 GYKDGKGNPGMGAP 13
Db 123 GYKDGKGNPGMGAP 135

RESULT 10
ID CA1F_HUMAN STANDARD; PRT: 1603 AA.
AC Q07092;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE COLLAGEN ALPHA 1(XVI) CHAIN PRECURSOR.
GN COL16A1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RA MEDLINE-92335339; PubMed-1631157;
RA Pan T.C., Zhang R.Z., Mattei M.-G., Timpl R., Chu M.-L.;
RT "Cloning and chromosomal location of human alpha 1(XVI) collagen.";
RL Proc. Natl. Acad. Sci. U.S.A. 89:6565-6569(1992).
RN [2]
RP SEQUENCE OF 418-1603 FROM N.A.
RC TISSUE-Placenta;
RA MEDLINE-93203161; PubMed-1284248;
RA Yamaguchi N., Kimura S., McBride O.W., Horl H., Yamada Y.,
RA Kanamori T., Yamakoshi H., Nagai Y.;
RT "Molecular cloning and partial characterization of a novel collagen
RT chain, alpha 1(XVI), consisting of repetitive collagenous domains and
RT cysteine-containing non-collagenous segments.";
RL J. Biochem. 112:856-863(1992).
CC -1- FUNCTION: THE NUMEROUS INTERRUPTIONS IN THE TRIPLE HELIX MAY MAKE
CC THIS MOLECULE EITHER ELASTIC OR FLEXIBLE.
CC -1- TISSUE SPECIFICITY: IN THE PLACENTA, WHERE IT IS FOUND IN THE
CC AMNION, A MEMBRANOUS TISSUE LINING THE AMNIOTIC CAVITY. WITHIN THE
CC COMPLEX NETWORK OF RETICULAR FIBERS, ALSO LOCATED TO A FIBROBLAST
CC LAYER BENEATH THIS DENSE LAYER, EXISTS IN TISSUES IN ASSOCIATION
CC WITH OTHER TYPES OF COLLAGEN.
CC -1- DEVELOPMENTAL STAGE: TRANSIENTLY ELEVATED EXPRESSION DURING
CC GESTATION, AND DECREASE AT TERM.
CC -1- DOMAIN: THIS SEQUENCE DEFINES EIGHTEEN DIFFERENT DOMAINS, NINE
CC TRIPLE-HELICAL DOMAINS (COL3 TO COL1) AND TEN NONTRIPLE-HELICAL
CC DOMAINS (NC10 TO NC1).
CC -1- PTM: PROLINES ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
CC -1- SIMILARITY: BELONGS TO THE FAMILY OF FIBRIL-ASSOCIATED COLLAGENS
CC WITH INTERRUPTED HELICES (FACIT).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M92642; AAA58427.1; -
CC DR EMBL: S57132; AAB25797.1; -
CC DR PIR: S23810; S23810.
CC MIM: 120326; -

```


FT DISULFID 1581 1634 OR 1637 (BY SIMILARITY).
 FT DISULFID 1593 1599 BY SIMILARITY.
 FT DISULFID 1656 1752 OR 1749 (BY SIMILARITY).
 FT DISULFID 1690 1749 OR 1752 (BY SIMILARITY).
 FT DISULFID 1702 1709 BY SIMILARITY.
 FT CARBOHYD 126 126 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 249 249 O-LINKED (GLYCOSAMINOGLYCAN) (IN FORM II)
 FT CARBOHYD 249 249 (POTENTIAL).
 FT VARSPPLIC 230 266 GEOGRGPGGPPGVPSTGAKGTITIGEGAPGPKGPK ->
 FT FT GDIGAPGPGPGPPGPPREFGSSIVAPRGHSGDKGVK (IN
 FT FT ISOFORM II).
 SQ SEQUENCE 1763 AA; 168526 MW; 304F52BEC06AAE0D CRC64;

Query Match 66.3%; Score 61; DB 1; Length 1763;
 Best Local Similarity 73.3%; Pred. No. 0.23;
 Matches 11; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 GYKGDGKNGPWCAP 15
 1: | | | | | | | | | |
 Db 760 GLPGMKNGPLPGAP 774

RESULT 13
 PSPD_BOVIN STANDARD; PRT; 369 AA.
 AC P35246;
 DT 01-FEB-1994 (Rel. 28, Created)
 DT 01-FEB-1994 (Rel. 28, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE PULMONARY SURFACTANT-ASSOCIATED PROTEIN D PRECURSOR (SP-D) (PSP-D).
 GN SFTPD OR SFTPD.
 OS Bos taurus (Bovine).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidea;
 OC Bovidae; Bovinae; Bos.
 NC NCBITaxID=9913;
 RN [1]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 208-247.
 RC TISSUE=Lung;
 RX MEDLINE=93170856; PubMed=8436402;
 RA Lim B.L., Lu J., Reid K.B.M.;
 RT "Structural similarity between bovine conglutinin and bovine lung
 RT surfactant protein D and demonstration of liver as a site of
 RT synthesis of conglutinin.";
 RT Immunology 78:159-165(1993).
 RL
 CC -1- FUNCTION: CONTRIBUTES TO THE LUNG'S DEFENSE AGAINST INHALED
 CC MICROORGANISMS. BINDS STRONGLY MALTOSE RESIDUES AND TO A LESSER
 CC EXTENT OTHER ALPHA-GLUCOSYL MOIETIES. IT COULD PARTICIPATE IN THE
 CC EXTRACELLULAR REORGANIZATION OR TURNOVER OF PULMONARY SURFACTANT.
 CC -1- SUBUNIT: OLIGOMERIC COMPLEX OF 4 SET OF HOMOTRIMERS.
 CC -1- SUBCELLULAR LOCATION: EXTRACELLULAR.
 CC -1- MISCELLANEOUS: PULMONARY SURFACTANT CONSISTS OF 90% LIPID AND 10%
 CC PROTEIN. THERE ARE 4 SURFACTANT ASSOCIATED PROTEIN: 2 COLLAGENOUS,
 CC CARBOHYDRATE-BINDING GLYCOPROTEINS (SP-A AND SP-D) AND 2 SMALL
 CC HYDROPHOBIC PROTEINS (SP-B AND SP-C).
 CC -1- SIMILARITY: CONTAINS A C-TYPE LECTIN FAMILY DOMAIN.
 CC -1- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC
 CC EMBL, X75911; CAAS3510.1; -.
 CC PIR, S33603; S33603.
 CC HSSP, P19999; 1MSB.
 CC InterPro: IPR000087; -.
 CC InterPro: IPR001304; -.
 CC Pfam: PF01391; Collagen; 3.

DR Pfam: PF00059; lectin_c; 1.
 DR PROSITE: PS00615; C-TYPE LECTIN_1; 1.
 DR PROSITE: PS00641; C-TYPE LECTIN_2; 1.
 KW Glycoprotein; Calcium; Surface film; Gaseous exchange; Hydroxylation;
 FT Signal; Lectin; collagen; Repeat; Coiled coil.
 FT SIGNAL 1 20
 FT CHAIN 1 369
 FT FT
 FT DOMAIN 46 216
 FT DOMAIN 217 248
 FT DOMAIN 273 369
 FT DISULFID 275 367 C-TYPE LECTIN (SHORT FORM).
 FT DISULFID 345 359 BY SIMILARITY.
 FT CARBOHYD 90 90 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT MOD_RES 78 78 HYDROXYLATION (BY SIMILARITY).
 FT MOD_RES 87 87 HYDROXYLATION (BY SIMILARITY).
 FT MOD_RES 96 96 HYDROXYLATION (BY SIMILARITY).
 FT MOD_RES 99 96 HYDROXYLATION (BY SIMILARITY).
 FT MOD_RES 165 165 HYDROXYLATION (BY SIMILARITY).
 FT MOD_RES 171 171 HYDROXYLATION (BY SIMILARITY).
 FT MOD_RES 171 171
 SQ SEQUENCE 369 AA; 37361 MW; 07D8B24E0AEB2E3 CRC64;

Query Match 65.2%; Score 60; DB 1; Length 369;
 Best Local Similarity 66.7%; Pred. No. 0.07;
 Matches 10; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 GYKGDGKNGPWCAP 15
 1: | | | | | | | | | |
 Db 157 GLKGERGAPGPGAP 171

RESULT 14
 CA44_RABIT STANDARD; PRT; 623 AA.
 AC P55787;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE COLLAGEN ALPHA 4(IV) CHAIN (FRAGMENT).
 GN COL4A4.
 OS Oryctolagus cuniculus (Rabbit).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
 NC NCBITaxID=9986;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Corneal endothelium;
 RX MEDLINE=93054733; PubMed=1429714;
 RA Kamagata Y., Mattei M.-G., Nimomiya Y.;
 RT "Isolation and sequencing of cDNAs and genomic DNAs encoding the
 RT alpha 4 chain of basement membrane collagen type IV and assignment of
 RT the gene to the distal long arm of human chromosome 2.";
 RL J. Biol. Chem. 267:23753-23758(1992).
 CC -1- FUNCTION: TYPE IV COLLAGEN IS THE MAJOR STRUCTURAL COMPONENT OF
 CC GLOMERULAR BASEMENT MEMBRANES (GBM), FORMING A 'CHICKEN-WIRE'
 CC MESHWORK TOGETHER WITH LAMININS, PROTEOLYCAN AND ENACTIN/
 CC NIIOGEN.
 CC -1- SUBUNIT: THERE ARE SIX TYPE IV COLLAGEN ISOFORMS, ALPHA 1(IV)-
 CC ALPHA 6(IV), EACH OF WHICH CAN FORM A TRIPLE HELIX STRUCTURE
 CC WITH 2 OTHER CHAINS TO GENERATE TYPE IV COLLAGEN NETWORK.
 CC -1- SUBCELLULAR LOCATION: CELL SURFACE (POTENTIAL).
 CC -1- DOMAIN: ALPHA CHAINS OF TYPE IV COLLAGEN HAVE A NONCOLLAGENOUS
 CC DOMAIN (NC1) AT THEIR C-TERMINUS, FREQUENT INTERRUPTIONS OF THE
 CC G-X-Y REPEATS IN THE LONG CENTRAL TRIPLE-HELICAL DOMAIN (WHICH MAY
 CC CAUSE FLEXIBILITY IN THE TRIPLE HELIX), AND A SHORT N-TERMINAL
 CC TRIPLE-HELICAL 7S DOMAIN.
 CC -1- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPLEPEPTIDE REPEATING
 CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
 CC -1- PTM: TYPE IV COLLAGENS CONTAIN NUMEROUS CYSTEINE RESIDUES WHICH
 CC ARE INVOLVED IN INTER- AND INTRAMOLECULAR DISULFIDE BONDING. 12 OF
 CC THESE, LOCATED IN THE NC1 DOMAIN, ARE CONSERVED IN ALL KNOWN TYPE
 CC IV COLLAGENS.

```

CC      UNIT G-X-Y ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
CC      DR EMBL; X70369; CAA49832.1; -.
CC      DR EMBL; M21354; AAA40942.1; -.
CC      DR PIR; A29905; A29905.
CC      DR PIR; S41067; S41067.
CC      DR InterPro; IPR000087; -.
CC      DR InterPro; IPR000085; -.
CC      DR InterPro; IPR001007; -.
CC      DR Pfam; PF01410; COLFR; 1.
CC      DR Pfam; PF01391; Collagen; 6.
CC      DR PROSITE; PS01208; WMFC; PARTIAL.
CC      KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
CC      Collagen; Glycoprotein.
CC      FT NON_TER 1 1
CC      FT CHAIN <1 375 COLLAGEN ALPHA 1(III) CHAIN.
CC      FT PROPEP 376 636 CARBOXYL-TERMINAL PROPEPTIDE.
CC      FT DOMAIN <1 368 TRIPLE-HELICAL REGION.
CC      FT DOMAIN 369 636 NONHELICAL REGION (C-TERMINAL).

```

50	SEQUENCE	636 AA;	62332 MW;	61A48159P01D01EB	CNC64;
	Query Match	65.2%;	Score 60;	DB 1;	Length 636;
	Best Local Similarity	71.4%;	Pred. NO. 0.12;		
	Matches 10;	Conservative	3;	Mismatches	1;
				Indels	0;
				Gaps	0;

DB	30	GVKGERGSPGRCGA	43
RESULT	16		
ID	CAL3-BOVIN	STANDARD:	PRT: 1049 AA.
AC	P04258;		
DT	20-MAR-1987 (Rel. 04, Created)		
DT	20-MAR-1987 (Rel. 04, Last sequence update)		
DT	15-JUL-1999 (Rel. 38, Last annotation update)		
DE	COLLAGEN ALPHA 1(III) CHAIN.		
GN	COL3A1.		
OS	Bos taurus (Bovine).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;		
OC	Bovidae; Bovinae; Bos.		
OX	NCBI_TaxID=9913;		
RN	[1]		
RP	SEQUENCE OF 1-242.		
RX	MEDLINE=80026026; PubMed=488906;		
RA	Fietzek P.P., Allmann H., Rautenberg J., Henkel W., Wachter E.,		
RA	Kuhn K.;		
RT	"The covalent structure of calf skin type III collagen. I. The amino		
RT	acid sequence of the amino terminal region of the alpha 1(III) chain		
RT	(positions 1-222).";		
RT	Hoppe-Seyler's Z. Physiol. Chem. 360: 809-820(1979).		
RN	[2]		
RP	SEQUENCE OF 243-422.		
RX	MEDLINE=80026027; PubMed=488907;		
RA	Dewes H., Fietzek P.P., Kuhn K.;		
RT	"The covalent structure of calf skin type III collagen. II. The amino		
RT	acid sequence of the cyanogen bromide peptide alpha 1(III)Cn1,8,10,2		
RT	(positions 223-402).";		
RT	Hoppe-Seyler's Z. Physiol. Chem. 360: 821-832(1979).		
RN	[3]		

RP SEQUENCE OF 423-571.
 RX MEDLINE=80026028; PubMed=488908;
 RA Benz H., Fietzek P.P., Kuhn K.;
 RT "The covalent structure of calf skin type III collagen. III. The
 RT amino acid sequence of the cyanogen bromide peptide alpha 1(III)CB4
 RT (positions 403-551).";
 RL Hoppe-Seyler's Z. Physiol. Chem. 360:833-840(1979).
 RN [4]
 RP SEQUENCE OF 572-808.
 RX MEDLINE=80026029; PubMed=488909;
 RA Lang H., Glanville R.W., Fietzek P.P., Kuhn K.;
 RT "The covalent structure of calf skin type III collagen. IV. The amino
 RT acid sequence of the cyanogen bromide peptide alpha 1(III)CB5
 RT (positions 552-788).";
 RL Hoppe-Seyler's Z. Physiol. Chem. 360:841-850(1979).
 RN [5]
 RP SEQUENCE OF 809-947.
 RX MEDLINE=80026030; PubMed=488910;
 RA Dewes H., Fietzek P.P., Kuhn K.;
 RT "The covalent structure of calf skin type III collagen. V. The amino
 RT acid sequence of the cyanogen bromide peptide alpha 1(III)CB9A
 RT (position 789-927).";
 RL Hoppe-Seyler's Z. Physiol. Chem. 360:851-860(1979).
 RN [6]
 RP SEQUENCE OF 948-1049.
 RX MEDLINE=80026031; PubMed=488911;
 RA Allmann H., Fietzek P.P., Glanville R.W., Kuhn K.;
 RT "The covalent structure of calf skin type III collagen. VI. The amino
 RT acid sequence of the carboxyterminal cyanogen bromide peptide alpha
 RT 1(III)CB9B (positions 928-1028).";
 RL Hoppe-Seyler's Z. Physiol. Chem. 360:861-868(1979).
 CC -1- FUNCTION: COLLAGEN TYPE III OCCURS IN MOST SOFT CONNECTIVE TISSUES
 CC ALONG WITH TYPE I COLLAGEN.
 CC -1- SUBUNIT: TRIMERS OF IDENTICAL ALPHA 1(III) CHAINS. THE CHAINS ARE
 CC LINKED TO EACH OTHER BY INTERCHAIN DISULFIDE BONDS. TRIMERS ARE
 CC ALSO CROSS-LINKED VIA HYDROXYLXINES.
 CC -1- PTH: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
 CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
 CC PIR: A02862; CGB07S.
 DR InterPro: IPR001007; -;
 DR InterPro: IPR001007; -;
 DR Pfam: PF01391; Collagen; 17.
 DR PROSITE: PS01208; WMFC; PARTIAL.
 KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
 KW Glycoprotein; Collagen.
 FT DOMAIN 1 14
 FT DOMAIN 15 1040 NONHELICAL REGION (N-TERMINAL).
 FT DOMAIN 1041 1049 TRIPLE-HELICAL REGION.
 FT MOD_RES 95 95 NONHELICAL REGION (C-TERMINAL).
 FT MOD_RES 107 107 HYDROXYLATION.
 FT MOD_RES 119 119 HYDROXYLATION.
 FT MOD_RES 938 938 HYDROXYLATION.
 FT MOD_RES 950 950 HYDROXYLATION.
 FT CARBOHYD 107 107 CROSS-LINK SITE.
 FT CARBOHYD 950 950 CROSS-LINK SITE.
 FT DISULFID 1040 1040 INTERCHAIN.
 FT DISULFID 1041 1041 INTERCHAIN.
 SQ SEQUENCE 1049 AA; 93651 MW; 8BEC33D1C66EC9A3 CRC64;

Query Match 65.28; Score 60; DB 1; Length 1049;
 Best Local Similarity 71.48; Pred. No. 0.19;
 Matches 10; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

OY 1 GVGDKGNPGMPGA 14
 DB 702 GVGKGRSGRGPFA 715
 AC P02461; STANDARD; PRT; 1466 AA.

DT 21-JUL-1986 (Rel. 01, Created)
 DT 01-JAN-1990 (Rel. 13, Last sequence update)
 DT 01-OCT-2000 (Rel. 40, Last annotation update)
 DE COLLAGEN ALPHA 1(III) CHAIN PRECURSOR.
 GN COL3A1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-Skin fibroblast;
 RX MEDLINE=89350838; PubMed=2764886;
 RA Ala-Kokko L., Kontusaari S., Baldwin C.T., Kuivaniemi H.,
 RA Prockop D.J.;
 RT "Structure of cDNA clones coding for the entire prepro alpha 1 (III)
 RT chain of human type III procollagen. Differences in protein structure
 RT from type I procollagen and conservation of codon preferences.";
 RL Biochem. J. 260:509-516(1989).
 RN [2]
 RP SEQUENCE OF 149-1225 FROM N.A.
 RX MEDLINE=89386015; PubMed=2780304;
 RA Janeczko R.A., Ramirez F.;
 RT "Nucleotide and amino acid sequences of the entire human alpha 1
 RT (III) collagen.";
 RL Nucleic Acids Res. 17:6742-6742(1989).
 RN [3]
 RP SEQUENCE OF 168-398.
 RX MEDLINE=77134724; PubMed=557335;
 RA Seyer J.M., Kang A.H.;
 RT "Covalent structure of collagen: amino acid sequence of cyanogen
 RT bromide peptides from the amino-terminal segment of type III collagen
 RT of human liver.";
 RL Biochemistry 16:1158-1164(1977).
 RN [4]
 RP REVISIONS.
 RA Seyer J.M.;
 RL Submitted (DEC-1977) to the PIR data bank.
 RN [5]
 RP SEQUENCE OF 399-727.
 RX MEDLINE=79000343; PubMed=687591;
 RA Seyer J.M., Kang A.H.;
 RT "Covalent structure of collagen: amino acid sequence of five
 RT consecutive CNBR peptides from type III collagen of human liver.";
 RL Biochemistry 17:3404-3411(1978).
 RN [6]
 RP SEQUENCE OF 728-964.
 RX MEDLINE=80198282; PubMed=6246925;
 RA Seyer J.M., Mainardi C., Kang A.H.;
 RT "Covalent structure of collagen: amino acid sequence of alpha 1
 RT (III)-CB5 from type III collagen of human liver.";
 RL Biochemistry 19:1583-1589(1980).
 RN [7]
 RP SEQUENCE OF 950-1466 FROM N.A.
 RX MEDLINE=88189827; PubMed=3357782;
 RA Mankoo B.S., Dalgleish R.;
 RT "Human pro alpha 1(III) collagen: cDNA sequence for the 3' end.";
 RL Nucleic Acids Res. 16:2337-2337(1988).
 RN [8]
 RP REVISION TO 1184.
 RX MEDLINE=89098346; PubMed=3211760;
 RA Molyneux K., Dalgleish R.;
 RT "Human type III collagen 'variant' is a cDNA cloning artefact.";
 RL Nucleic Acids Res. 16:11833-11833(1988).
 RN [9]
 RP SEQUENCE OF 1065-1466 FROM N.A.
 RX MEDLINE=85087944; PubMed=6096827;
 RA Loidl H.R., Brinker J.M., May M., Pihlajaniemi T., Morrow S.,
 RA Rosenbloom J., Myers J.C.;
 RT "Molecular cloning and carboxyl-propeptide analysis of human type III
 RT procollagen.";
 RL Nucleic Acids Res. 12:9383-9394(1984).
 RN [10]

RP SEQUENCE OF 965-1200.
 RX MEDLINE-81208139; PubMed-7016180;
 RA Seyer J.M., Kang A.H.;
 RT "Covalent structure of collagen: amino acid sequence of alpha
 RT 1(III)-CB9 from type III collagen of human liver.";
 RL Biochemistry 20:2621-2627(1981).
 RN [11]
 RP SEQUENCE OF 1176-1466 FROM N.A.
 RX MEDLINE-85157600; PubMed-2579949;
 RA Chu M.-L., Weil D., de Wet W.J., Bernard M.P., Sippola M., Ramirez F.;
 RT "Isolation of cDNA and genomic clones encoding human pro-alpha 1
 RT (III) collagen. Partial characterization of the 3' end region of the
 RT gene.";
 RL J. Biol. Chem. 260:4357-4363(1985).
 RN [12]
 RP SEQUENCE OF 1161-1200 FROM N.A.
 RX MEDLINE-86187804; PubMed-3754462;
 RA Miskulin M., Dalgleish R., Kluge-Beckerman B., Rennard S.I.,
 RA Tolstoshev P., Brantly M., Crystal R.G.;
 RT "Human type III collagen gene expression is coordinately modulated
 RT with the type I collagen genes during fibroblast growth.";
 RL Biochemistry 25:1408-1413(1986).
 RN [13]
 RP SEQUENCE OF 1-170 FROM N.A.
 RX TISSUE-Placenta;
 RC MEDLINE-88303360; PubMed-3405773;
 RA Toman D., Rices G., de Crombrughe B.;
 RT "Nucleotide sequence of a cDNA coding for the amino-terminal region
 RT of human proalpha 1(III) collagen.";
 RL Nucleic Acids Res. 16:7201-7201(1988).
 RN [14]
 RP SEQUENCE OF 1-176 FROM N.A.
 RX MEDLINE-89378752; PubMed-2777083;
 RA Benson-Chanda V., Su M.W., Weil D., Chu M.L., Ramirez F.;
 RT "Cloning and analysis of the 5' portion of the human type-III
 RT procollagen gene (COL3A1).";
 RL Gene 78:255-265(1989).
 RN [15]
 RP REVIEW ON VARIANTS.
 RX MEDLINE-97255959; PubMed-9101290;
 RA Kivianlehti H., Tromp G., Prockop D.J.;
 RT "Mutations in fibrillar collagens (types I, II, III, and XI), fibril-
 RT associated collagen (type IX), and network-forming collagen (type X)
 RT cause a spectrum of diseases of bone, cartilage, and blood vessels.";
 RL Hum. Mutat. 9:300-315(1997).
 RN [16]
 RP VARIANT AORTIC ANEURYSM ARG-303;
 RX MEDLINE-93293988; PubMed-8514866;
 RA Tromp G., Wu Y., Prockop D.J., Madhathari S.L., Kleiner C.,
 RA Earley J.J., Zhuang J., Noerregaard O., Darling R.C., Abbott W.M.,
 RA Cole C.W., Jaakkola P., Ryyanen M., Pearce W.H., Yao J.S.T.,
 RA Majamaa K., Smulders V.V., Gatalica Z., Ferrelli R.E., Jimenez S.A.,
 RA Jackson C.E., Michaels V.V., Kaye M., Kivianlehti H.;
 RT "Sequencing of cDNA from 50 unrelated patients reveals that mutations
 RT in the triple-helical domain of type III procollagen are an
 RT infrequent cause of aortic aneurysms.";
 RL J. Clin. Invest. 91:2539-2545(1993).
 RN [17]
 RP VARIANT THR-689
 RX MEDLINE-91045136; PubMed-2235526;
 RA Zafarullah K., Kleiner C., Tromp G., Kivianlehti H., Kontusaari S.,
 RA Wu Y., Ganguly A., Prockop D.J.;
 RT "G to A polymorphism in exon 31 of the COL3A1 gene.";
 RL Nucleic Acids Res. 18:6180-6180(1990).
 RN [18]
 RP VARIANT AORTIC ANEURYSM ARG-786
 RX MEDLINE-91056145; PubMed-2243125;
 RA Kontusaari S., Tromp G., Kivianlehti H., Romanic A.M., Prockop D.J.;
 RT "A mutation in the gene for type III procollagen (COL3A1) in a family
 RT with aortic aneurysms.";
 RL J. Clin. Invest. 86:1465-1473(1990).
 RN [19]
 RP VARIANT EDS-IV ARG-828.

RX MEDLINE-94016385; PubMed-8411057;
 RA Richards A., Narcisi P., Lloyd J., Ferguson C., Pope F.M.;
 RT "The substitution of glycine 661 by arginine in type III collagen
 RT produces mutant molecules with different thermal stabilities and
 RT causes Ehlers-Danlos syndrome type IV.";
 RL J. Med. Genet. 30:690-693(1993).
 RN [20]
 RP VARIANT EDS-IV SER-957.
 RX MEDLINE-89109135; PubMed-2492273;
 RA Tromp G., Kivianlehti H., Shikata H., Prockop D.J.;
 RT "A single base mutation that substitutes serine for glycine 790 of
 RT the alpha 1 (III) chain of type III procollagen exposes an arginine
 RT and causes Ehlers-Danlos syndrome IV.";
 RL J. Biol. Chem. 264:1349-1352(1989).
 RN [21]
 RP VARIANT EDS-IV VAL-960.
 RX MEDLINE-95268429; PubMed-7749417;
 RA Tromp G., de Paeppe A., Nuytink L., Madhathari S., Kivianlehti H.;
 RT "Substitution of valine for glycine 793 in type III procollagen in
 RT Ehlers-Danlos syndrome type IV.";
 RL Hum. Mutat. 5:179-181(1995).
 RN [22]
 RP VARIANT EDS-IV GLU-1014.
 RX MEDLINE-92316511; PubMed-1352273;
 RA Richards A.J., Ward P.N., Narcisi P., Nicholls A.C., Lloyd J.C.,
 RA Pope F.M.;
 RT "A single base mutation in the gene for type III collagen (COL3A1)
 RT converts glycine 847 to glutamic acid in a family with Ehlers-Danlos
 RT syndrome type IV. An unaffected family member is mosaic for the
 RT mutation.";
 RL Hum. Genet. 89:414-418(1992).
 RN [23]
 RP VARIANT EDS-IV ASP-1050.
 RX MEDLINE-90037070; PubMed-2808425;
 RA Tromp G., Kivianlehti H., Stolle C., Pope F.M., Prockop D.J.;
 RT "Single base mutation in the type III procollagen gene that converts
 RT the codon for glycine 883 to aspartate in a mild variant of
 RT Ehlers-Danlos syndrome IV.";
 RL J. Biol. Chem. 264:19313-19317(1989).
 RN [24]
 RP VARIANT EDS-IV VAL-1077.
 RX MEDLINE-91374480; PubMed-1895316;
 RA Richards A.J., Lloyd J.C., Ward P.N., de Paeppe A., Narcisi P.,
 RA Pope F.M.;
 RT "Characterisation of a glycine to valine substitution at amino acid
 RT position 910 of the triple helical region of type III collagen in a
 RT patient with Ehlers-Danlos syndrome type IV.";
 RL J. Med. Genet. 28:458-463(1991).
 RN [25]
 RP VARIANT EDS-IV GLU-1173.
 RX MEDLINE-9302543; PubMed-1357232;
 RA Johnson P.H., Richards A.J., Pope F.M., Hopkinson D.A.;

Query Match 65.2%; Score 60; DB 1; Length 1466;
 Best Local Similarity 78.6%; Pred. No. 0.27;
 Matches 11; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GYKGDKNPCMGCA 14
 DB 738 GPKDKGEPGPGCA 751

RESULT 18
 CA34_BOVIN
 ID CA34_BOVIN STANDARD; PRT; 471 AA.
 AC 028084;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE COLLAGEN ALPHA 3(IV) CHAIN (FRAGMENT).
 GN COL4A3.
 OS Bos taurus (Bovine).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 OC Bovidae; Bovinae; Bos.
 OC NCB1_TaxID=9913;
 [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Lens;
 RA MEDLINE=91093146; PubMed=1985905;
 RT Morrison K.E., Germino G.G., Reeders S.T.;
 "Use of the polymerase chain reaction to clone and sequence a cDNA
 encoding the bovine alpha 3 chain of type IV collagen.";
 RL J. Biol. Chem. 266:34-39(1991).
 [2]
 RP SEQUENCE OF 227-258.
 RC TISSUE=Kidney;
 RA MEDLINE=90202779; PubMed=2318822;
 RT Gunwar S., Saus J., Noelken M.E., Hudson B.G.;
 "Glomerular basement membrane. Identification of a fourth chain,
 alpha 4, of type IV collagen.";
 RL J. Biol. Chem. 265:5466-5469(1990).
 [3]
 RP SEQUENCE OF 227-254.
 RC MEDLINE=88330844; PubMed=3417661;
 RA Saus J., Wieslander J., Langeveld J.P.M., Quinones S., Hudson B.G.;
 "Identification of the goodpasture antigen as the alpha 3(IV) chain
 of collagen IV.";
 RL J. Biol. Chem. 263:13374-13380(1988).
 [4]
 RP SEQUENCE OF 227-244.
 RC MEDLINE=87222419; PubMed=2438283;
 RA Butkowski R.J., Langeveld J.P.M., Wieslander J., Hamilton J.,
 Hudson B.G.;
 "Localization of the Goodpasture epitope to a novel chain of basement
 membrane collagen.";
 RL J. Biol. Chem. 262:7874-7877(1987).
 CC -1- FUNCTION: TYPE IV COLLAGEN IS THE MAJOR STRUCTURAL COMPONENT OF
 GLOMERULAR BASEMENT MEMBRANES (GBM), FORMING A 'CHICKEN-WIRE'
 MESHWORK TOGETHER WITH LAMININS, PROTEGLYCANS AND ENACTIN/
 NIDOGEM.
 CC -1- SUBUNIT: THERE ARE SIX TYPE IV COLLAGEN ISOFORMS, ALPHA 1(IV)-
 ALPHA 6(IV), EACH OF WHICH CAN FORM A TRIPLE HELIX STRUCTURE
 WITH 2 OTHER CHAINS TO GENERATE TYPE IV COLLAGEN NETWORK.
 CC -1- SUBCELLULAR LOCATION: CELL SURFACE (POTENTIAL).
 CC -1- DOMAIN: ALPHA CHAINS OF TYPE IV COLLAGEN HAVE A NONCOLLAGENOUS
 DOMAIN (NC1) AT THEIR C-TERMINUS, FREQUENT INTERRUPTIONS OF THE
 G-X-Y REPEATS IN THE LONG CENTRAL TRIPLE-HELICAL DOMAIN (WHICH MAY
 CAUSE FLEXIBILITY IN THE TRIPLE HELIX), AND A SHORT N-TERMINAL
 TRIPLE-HELICAL 7S DOMAIN.
 CC -1- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPETIDE REPEATING
 UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
 CC -1- PTM: TYPE IV COLLAGENS CONTAIN NUMEROUS CYSTEINE RESIDUES WHICH
 ARE INVOLVED IN INTER- AND INTRAMOLECULAR DISULFIDE BONDING. 12 OF
 THESE, LOCATED IN THE NC1 DOMAIN, ARE CONSERVED IN ALL KNOWN TYPE
 IV COLLAGENS.
 CC -1- SIMILARITY: TO OTHER TYPE IV COLLAGENS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: M63139; AAA62708.1; -;
 DR InterPro: IPR000087; -;
 DR InterPro: IPR001442; -;
 DR Pfam: PF01413; C4; 2;
 DR Pfam: PF01391; Collagen; 4;
 KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
 KW Glycoprotein; Basement membrane; Collagen; Cell adhesion.
 FT NON_TER 1 1
 FT DOMAIN <1 238 TRIPLE-HELICAL REGION.
 FT DOMAIN 239 471 NONHELICAL REGION (NC1).

FT SITE 106 108 CELL ATTACHMENT SITE (POTENTIAL).
 FT MOD_RES 232 232 HYDROXYLATION.
 FT MOD_RES 238 238 HYDROXYLATION.
 FT DISULFID 261 352 OR 349 (BY SIMILARITY).
 FT DISULFID 294 349 OR 352 (BY SIMILARITY).
 FT DISULFID 306 312 BY SIMILARITY.
 FT DISULFID 371 466 OR 463 (BY SIMILARITY).
 FT DISULFID 405 463 OR 466 (BY SIMILARITY).
 FT DISULFID 417 423 BY SIMILARITY.
 FT CONFLICT 253 253 S -> Y (IN REF. 3).
 SQ SEQUENCE 471 AA; 47585 MW; C03B6F14E7008DE CRC64;
 Query Match 64.1%; Score 59; DB 1; Length 471;
 Best local Similarity 66.7%; Pred. No. 0.12;
 Matches 10; Conservative 2; Mismatches 3; Indels 0; Gaps 0;
 QY 1 GYKGDKNPGWPGAP 15
 DB 116 GYKGDKNPGWPGAP 130
 ID CAL4_CAEEL STANDARD; PRT; 1758 AA.
 AC P17139;
 DT 01-AUG-1990 (Rel. 15, Created)
 DT 01-MAR-1992 (Rel. 21, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE COLLAGEN ALPHA 1(IV) CHAIN PRECURSOR.
 GN EMB-9 OR CLB-2 OR K04H4.1.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nemata; Rhabditida; Rhabditoidea;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 OC NCB1_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RA MEDLINE=91141582; PubMed=1996137;
 RA Guo X., Johnson J.J., Kramer J.M.;
 "Embryonic lethality caused by mutations in basement membrane
 collagen of C. elegans";
 RL Nature 349:707-709(1991).
 [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RA MEDLINE=94150718; PubMed=7906398;
 RA Wilson R., Almscough R., Anderson K., Baynes C., Berks M.,
 RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
 RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fraser A.,
 RA Fulton L., Gardner A., Green P., Hawkins T., Hillier L., Jier M.,
 RA Johnston L., Jones M., Kerhaw J., Kirsten J., Laister N.,
 RA Latreille P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M.,
 RA Parsons J., Percy C., Rifkin L., Roopra A., Saunders D., Showkhen R.,
 RA Sims M., Smalton N., Smith A., Smith M., Sonhammer E., Staden R.,
 RA Sulston J., Thierley-Mieg J., Thomas K., Vaudin M., Vaughan K.,
 RA Waterson P., Watson A., Welstock L., Wilkinson-Spoat J.,
 RA Woldman P.;
 "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
 elegans";
 RL Nature 368:32-38(1994).
 [3]
 RP SEQUENCE OF 1446-1758 FROM N.A.
 RC STRAIN-BRISTOL N2;
 RA MEDLINE=90008929; PubMed=2793871;
 RA Guo X., Kramer J.M.;
 "The two Caenorhabditis elegans basement membrane (type IV) collagen
 genes are located on separate chromosomes";
 RL J. Biol. Chem. 264:17574-17582(1989).
 CC -1- FUNCTION: COLLAGEN TYPE IV IS SPECIFIC FOR BASEMENT MEMBRANES.
 CC -1- SUBUNIT: TRIMERS OF TWO ALPHA 1(IV) AND ONE ALPHA 2(IV) CHAIN.
 CC TYPE IV COLLAGEN FORMS A MESH-LIKE NETWORK LINKED THROUGH
 CC INTERMOLECULAR INTERACTIONS BETWEEN 7S DOMAINS AND BETWEEN NC1

"Complete primary structure of two variant forms of human type XVIII collagen and tissue-specific differences in the expression of the corresponding transcripts.";
 Matrix Biol. 16:319-328(1998).
 [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-20289799; PubMed-10830953;
 RA Hattori M., Fujiyama A., Taylor T.D., Watanabe H., Yada T.,
 RA Park H.-S., Toyoda A., Ishii K., Totoki Y., Taudien S., Blechschmidt K., Polley A.,
 RA Ohki M., Takagi T., Sakaki Y.,
 RA Menzel U., Delabber J., Kumpf K., Lehmann R., Patterson D.,
 RA Reinwald K., Kump A., Schillabel M., Schudy A., Zimmermann W.,
 RA Rosenthal A., Kudo J., Shibuya K., Kawasaki K., Asakawa S.,
 RA Shintani A., Sasaki T., Nagamine K., Mitsuyma S., Antonarakis S.E.,
 RA Minoshima S., Shimizu N., Nordstiek G., Hornischer K., Brandt P.,
 RA Scharfe M., Schoen O., Desario A., Reichelt J., Kauer G., Bloecher H.,
 RA Ramser J., Beck A., Klages S., Hennig S., Rieselmann L., Dandand E.,
 RA Weinmeyer S., Borzym K., Gardiner K., Mizetic D., Francis F.,
 RA Lehnach H., Reinhardt R., Yaspo M.-L.,
 RA "The DNA sequence of human chromosome 21.";
 RA Nature 405:311-319(2000).
 RL [3]
 RP SEQUENCE OF 834-1516 FROM N.A.
 RX MEDLINE-94245237; PubMed-8188291;
 RA Oh S.P., Warman M.L., Seldin M.F., Cheng S., Knoll J.H., Timmons S.,
 RA Olsen B.R.,
 RT "Cloning of cDNA and genomic DNA encoding human type XVIII collagen
 RT and localization of the alpha 1(XVIII) collagen gene to mouse
 RT chromosome 10 and human chromosome 21.";
 RL Genomics 19:494-499(1994).
 RN [4]
 RP SEQUENCE OF 1334-1516 FROM N.A.
 RC TISSUE-placenta;
 RA Zhi-Yong H., Biao L., Wei-Jie Z., Xiang-Fu W.,
 RT "Cloning and expression of human endostatin gene in Escherichia
 RT coli.";
 RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
 RN [5]
 RP INVOLVEMENT IN KNOBLOCH SYNDROME.
 RX MEDLINE-20400145; PubMed-10942434;
 RA Serite A.L., Sossi V., Camargo A.A., Zatz M., Brahe C.,
 RA Passos-Bueno M.R.,
 RT "Collagen XVIII, containing an endogenous inhibitor of angiogenesis
 RT and tumor growth, plays a critical role in the maintenance of retinal
 RT structure and in neural tube closure.";
 RL Hum. Mol. Genet. 9:2051-2058(2000).
 CC -1- FUNCTION: COL18A1 PROBABLY PLAYS A MAJOR ROLE IN DETERMINING THE
 CC RETINAL STRUCTURE AS WELL AS IN THE CLOSURE OF THE NEURAL TUBE.
 CC -1- FUNCTION: ENDOSTATIN POTENTIALLY INHIBITS ENDOTHELIAL CELL
 CC PROLIFERATION AND ANGIOGENESIS. MAY INHIBIT ANGIOGENESIS BY
 CC BINDING TO THE HEPARAN SULPHATE PROTEOGLYCANS INVOLVED IN GROWTH
 CC FACTOR SIGNALING (BY SIMILARITY).
 CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A SHORT FORM/NC1-303 AND A LONG
 CC FORM/NC-493 (SHOWN HERE); ARE PRODUCED BY ALTERNATIVE SPLICING.
 CC -1- TISSUE SPECIFICITY: PRESENT IN MULTIPLE ORGANS WITH HIGHEST LEVELS
 CC IN LIVER, LUNG AND KIDNEY.
 CC -1- PTM: PROLINES ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
 CC -1- DISEASE: DEFECTS IN COL18A1 ARE A CAUSE OF KNOBLOCH SYNDROME (KS);
 CC AN AUTOSOMAL RECESSIVE DISORDER DEFINED BY THE OCCURRENCE OF HIGH
 CC MYOPIA, VITREORETINAL DEGENERATION WITH RETINAL DETACHMENT,
 CC MACULAR ABNORMALITIES AND OCCIPITAL ENCEPHALOCELE.
 CC -1- SIMILARITY: BELONGS TO THE MULTIPLEXIN FAMILY OF COLLAGENS.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL; AF018081; AAC39658.1; -

DR EMBL; AF018082; AAC39659.1; -
 DR EMBL; AL163302; CAB90482.1; -
 DR EMBL; I22548; AAA51864.1; -
 DR EMBL; AF184060; AAF01310.1; ALT_INIT.
 DR GlycoSuiteDB; P39060; -
 DR MIM; 120328; -
 DR MIM; 267750; -
 DR InterPro: IPR000087; -
 DR Pfam: PF01391; Collagen; 6.
 KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
 KW Cell adhesion; Collagen; Glycoprotein; Signal; Alternative splicing.
 FT SIGNAL 1 23
 FT CHAIN 1 1516
 FT CHAIN 1334 1516
 FT CHAIN 24 516
 FT DOMAIN 517 550
 FT DOMAIN 531 560
 FT DOMAIN 551 640
 FT DOMAIN 641 664
 FT DOMAIN 665 786
 FT DOMAIN 787 809
 FT DOMAIN 810 892
 FT DOMAIN 893 906
 FT DOMAIN 907 948
 FT DOMAIN 949 961
 FT DOMAIN 962 1034
 FT DOMAIN 1035 1044
 FT DOMAIN 1045 1077
 FT DOMAIN 1078 1089
 FT DOMAIN 1090 1111
 FT DOMAIN 1112 1118
 FT DOMAIN 1119 1173
 FT DOMAIN 1174 1186
 FT DOMAIN 1187 1204
 FT DOMAIN 1205 1516
 FT CARBOHYD 68 68
 FT CARBOHYD 129 129
 FT CARBOHYD 164 164
 FT CARBOHYD 691 691
 FT DISULFID 1366 1506
 FT DISULFID 1468 1498
 FT SITE 1095 1097
 FT SITE 180 180
 FT VARSPLIC 1 215
 FT VARSPLIC 181 215
 FT CONFLICT 428 428
 FT CONFLICT 841 841
 FT CONFLICT 877 877
 FT CONFLICT 886 886
 FT CONFLICT 912 912
 FT CONFLICT 933 933
 FT CONFLICT 975 975
 FT CONFLICT 1064 1064
 FT CONFLICT 1084 1084
 FT CONFLICT 1108 1120
 FT CONFLICT 1123 1123
 FT CONFLICT 1126 1126
 FT CONFLICT 1126 1126
 FT CONFLICT 1206 1306
 FT CONFLICT 1304 1304
 FT CONFLICT 1314 1314
 FT CONFLICT 1314 1314
 FT CONFLICT 1324 1324
 FT CONFLICT 1437 1437
 FT CONFLICT 1443 1443
 FT CONFLICT 1483 1483
 FT CONFLICT 1483 1483
 FT SEQUENCE 1516 AA; 153840 MW; 3c70f29a4476ee76 CRC64;
 F -> S (IN REF. 2).
 I -> V (IN REF. 2).
 V -> L (IN REF. 3).
 P -> R (IN REF. 3).
 P -> R (IN REF. 3).
 R -> L (IN REF. 3).
 P -> L (IN REF. 3).
 A -> P (IN REF. 3).
 L -> K (IN REF. 3).
 P -> A (IN REF. 3).
 P -> A (IN REF. 3).
 P -> A (IN REF. 3).
 P -> P (IN REF. 3).
 G -> GO (IN REF. 3).
 R -> G (IN REF. 3).
 A -> G (IN REF. 3).
 LR -> CG (IN REF. 3).
 D -> N (IN REF. 4).
 R -> T (IN REF. 4).
 S -> Y (IN REF. 4).
 CELL ATTACHMENT SITE (POTENTIAL).
 MISSING (IN SHORT ISOFORM).
 HTTEAGTLPAIPSPSPSGRPAPVLTGVSPPSS -> MA
 PCPCWMPRRRLDVLAVLLGVRRAASAP (IN
 SHORT ISOFORM).

Query Match 63.0%; Score 58; DB 1; Length 1516;
 Best Local Similarity 66.7%; Pred. No. 0.54;
 Matches 10; Conservative 1; Mismatches 4; Indels 0; Gaps 0;
 QY 1 GYGDKGNDGPGAP 15

DB 517 GLKGGKGGPGVP 531

RESULT 22

CA24_CAEEL STANDARD: PRT: 1758 AA.

ID CA24_CAEEL

AC P17140;

DT 01-AUG-1990 (Rel. 15, Created)

DT 01-OCT-1994 (Rel. 30, Last sequence update)

DT 30-MAY-2000 (Rel. 39, Last annotation update)

DE COLLAGEN ALPHA 2(IV) CHAIN PRECURSOR.

GN LET-2 OR CLB-1.

OS Caenorhabditis elegans.

OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea; Rhabditidae; Pelodierinae; Caenorhabditis.

OX NCBI_Taxid=6239;

ON (1)

RP SEQUENCE FROM N.A.

RC STRAIN-BRISTOL N2;

RX MEDLINE-94012964; PubMed-7691828;

RA Sibley M.H., Johnson J.J., Mello C.C., Kramer J.M.;

RT "Genetic identification, sequence, and alternative splicing of the Caenorhabditis elegans alpha 2(IV) collagen gene.";

RL J. Biol. Chem. 123:255-264(1993).

RM (2)

RP PRELIMINARY SEQUENCE OF 1495-1758 FROM N.A.

RC STRAIN-BRISTOL N2;

RX MEDLINE-90008929; PubMed-2793871;

RA Guo X., Kramer J.M.;

RT "The two Caenorhabditis elegans basement membrane (type IV) collagen genes are located on separate chromosomes.";

RL J. Biol. Chem. 264:17574-17582(1989).

RN [3]

RP VARIANTS

RX MEDLINE-94320591; PubMed-80455258;

RA Sibley M.H., Graham P.L., von Mendel N., Kramer J.M.;

RT "Mutations in the alpha 2(IV) basement membrane collagen gene of Caenorhabditis elegans produce phenotypes of differing severities.";

RL EMBO J. 13:3278-3285(1994).

CC -1- FUNCTION: COLLAGEN TYPE IV IS SPECIFIC FOR BASEMENT MEMBRANES.

CC -1- SUBUNIT: TRIMERS OF TWO ALPHA 1(IV) AND ONE ALPHA 2(IV) CHAIN.

CC TYPE IV COLLAGEN FORMS A MESH-LIKE NETWORK LINKED THROUGH INTERMOLECULAR INTERACTIONS BETWEEN 7S DOMAINS AND BETWEEN NC1 DOMAINS.

CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS: I (SHOWN HERE) AND II; ARE PRODUCED BY ALTERNATIVE SPLICING.

CC -1- TISSUE SPECIFICITY: FORM I IS PREDOMINANT IN EMBRYOS AND FORM II IS PREDOMINANT IN THE LARVAE AND ADULTS.

CC -1- DOMAIN: ALPHA CHAINS OF TYPE IV COLLAGEN HAVE A NONCOLLAGENOUS DOMAIN (NC1) AT THEIR C-TERMINUS, FREQUENT INTERRUPTIONS OF THE G-X-Y REPEATS IN THE LONG CENTRAL TRIPLE-HELICAL DOMAIN (WHICH MAY CAUSE FLEXIBILITY IN THE TRIPLE HELIX), AND A SHORT N-TERMINAL TRIPLE-HELICAL 7S DOMAIN.

CC -1- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.

CC -1- PTM: TYPE IV COLLAGENS CONTAIN NUMEROUS CYSTEINE RESIDUES WHICH ARE INVOLVED IN INTER- AND INTRAMOLECULAR DISULFIDE BONDING. 12 OF THESE, LOCATED IN THE NC1 DOMAIN, ARE CONSERVED IN ALL KNOWN TYPE IV COLLAGENS.

CC -1- DISEASE: MUTATIONS IN LET-2 ARE GENERALLY EMBRYONIC LETHAL.

CC -----

CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----

DR EMBL: Z22964; CAAB0536.1; -

DR EMBL: Z22964; CAAB0537.1; -

DR EMBL: J05066; AAA27989.1; -

DR EMBL: U22327; AAA64312.1; ALT_SEQ.

DR PIR: A34476; A34476.

DR InterPro: IPR000087; -

DR InterPro: IPR001442; -

DR Pfam: PF01413; C4; 2.

DR Pfam: PF01391; Collagen; 23.

KW Hydroxylation; Connective tissue; Basement membrane; Repeat; Collagen; Alternative splicing; Glycoprotein; Signal.

KW P01391; Collagen; 23.

FT CHAIN 1 26

FT SIGNAL 27 1758

FT DOMAIN 27 42

FT DOMAIN 42 1527

FT DOMAIN 1528 1758

FT DISULFID 1546 1635

FT DISULFID 1579 1632

FT DISULFID 1591 1597

FT DISULFID 1654 1750

FT DISULFID 1688 1747

FT DISULFID 1700 1707

FT CARBOHYD 248 248

FT VARSPLIC 229 264

FT VARIANT 48 48

FT VARIANT 366 366

FT VARIANT 570 570

FT VARIANT 588 588

FT VARIANT 597 597

FT VARIANT 690 690

FT VARIANT 690 690

FT VARIANT 737 737

FT VARIANT 877 877

FT VARIANT 904 904

FT VARIANT 1003 1003

FT VARIANT 1125 1125

FT VARIANT 1152 1152

FT VARIANT 1286 1286

FT CONFLICT 1682 1682

SO SEQUENCE 1758 AA; 167750 MW; 97EE3F3DBB2D2AC5 CRC64;

Query Match 63.0%; Score 58; DB 1; Length 1758;

Best Local Similarity 73.3%; Pred. No. 0.62;

Matches 11; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 GVGKGGKGGPGVP 15

DB 416 GPRGDKGGGPGAP 430

RESULT 23

CC34_CAEEL STANDARD: PRT: 298 AA.

ID CC34_CAEEL

AC P34687;

DT 01-FEB-1994 (Rel. 28, Created)

DT 01-FEB-1994 (Rel. 28, Last sequence update)

DT 01-JUN-1994 (Rel. 29, Last annotation update)

DE CUTICLE COLLAGEN 34.

GN COL-34.

OS Caenorhabditis elegans.

OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea; Rhabditidae; Pelodierinae; Caenorhabditis.

OX NCBI_Taxid=6239;

ON (1)

RP SEQUENCE FROM N.A.

RC STRAIN-BRISTOL N2;

RX MEDLINE-93013043; PubMed-1398138;

RA Bird D.M.;

RT "Sequence comparison of the Caenorhabditis elegans dpy-13 and col-34 gene 120:261-266(1992).

CC -1- FUNCTION: NEMATODE CUTICLES ARE COMPOSED LARGELY OF COLLAGEN-LIKE PROTEINS. THE CUTICLE FUNCTIONS BOTH AS AN EXOSKELETON AND AS A

Query Match	62.0%	Score 57;	DB 1;	Length 298;
Best Local Similarity	73.3%	Pred. NO. 0.16;		
Matches 11; Conservative	0;	Mismatches 4;	Indels 0;	Gaps 0;
OY	1	GVKGDKNPMPGAP	15	
1b	184	GPKEGKPPGAPGAP	198	

RC	STRAIN=C57BL;
RX	MEDLINE=91274355; PubMed=2054384;
RA	MeSeaneanta M., Roman D., de Crombrughe B., Viorio E.;
RT	"Specific hybridization probes for mouse type I, II, III and IX collagen mRNAs.";
RL	Biochim. Biophys. Acta 1089:241-243(1991).
CC	- FUNCTION: COLLAGEN TYPE III OCCURS IN MOST SOFT CONNECTIVE TISSUES
CC	ALONG WITH TYPE I COLLAGEN.
CC	- SUBUNIT: TRIMERS OF IDENTICAL ALPHA 1(III) CHAINS. THE CHAINS ARE LINKED TO EACH OTHER BY INTERCHAIN DISULFIDE BONDS. TRIMERS ARE ALSO CROSS-LINKED VIA HYDROXYLINES.
CC	- PWM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
CC	- LINKED GLICAN CONSISTS OF GIC-GAL DISACCHARIDE (BY SIMILARITY).
CC	- SIMILARITY: CONTAINS 1 WVEC DOMAIN.
CC	-----
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@sib-sib.ch).
CC	-----
DR	EMBL; X52046; CAA36279.1; -;
DR	EMBL; M18933; AAA37338.1; -;
DR	EMBL; K03037; -; NOT_ANNOTATED_CDS.
DR	EMBL; X57983; CAA41048.1; -;
DR	PIR; A22287; A22287.
DR	PIR; A27353; A27353.
DR	PIR; S1673; S1673.
DR	MGI; MGI:88453; Col3a1.
DR	InterPro; IPRO000087; -;
DR	InterPro; IPRO00085; -;
DR	InterPro; IPRO01007; -;
DR	Pfam; PF01410; COL1; 1.
DR	Pfam; PF01391; Collagen; 17.
DR	PROSITE; PS01208; WVEC; 1.
KW	Extracellular matrix; Connective tissue; Repeat; Hydroxylation; Glycoprotein; Collagen; Signal.
FT	SIGNAL 1 23
FT	PROPEP 24 154
FT	CHAIN 155 1203
FT	PROPER 1204 1464
FT	DOMAIN 31 90
FT	DOMAIN 155 169
FT	DOMAIN 170 1195
FT	DOMAIN 1196 1464
FT	CARBOHYD 262 262
FT	MOD_RES 262 262
FT	MOD_RES 283 283
FT	MOD_RES 859 859
FT	MOD_RES 976 976
FT	MOD_RES 1093 1093
FT	MOD_RES 1105 1105
FT	DISULFD 1195 1195
FT	DISULFD 1196 1196
SQ	SEQUENCE 1464 AA; 138944 MW; 2104EC27A8B6090B CRC64;
QY	Query Match 60.9%; Score 56; DB 1; Length 1464; Best Local Similarity 69.2%; Pred. No. 1; Matches 9; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
ID	MSRE_MOUSE STANDARD; PRT; 458 AA. P30204;
RESULT	25
DB	857 GVKGKGNPGMPG 13 1111::1::11 11 GVKGKGNPGMPG 869

DT 01-APR-1993 (Rel. 25, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE MACROPHAGE SCAVENGER RECEPTOR TYPES I AND II (MACROPHAGE ACETYLATED
 LDL RECEPTOR I AND II).
 GN MSRI OR SCVR.
 OC Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 NC NCBL_TaxID=10090;
 RP [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=93359822; PubMed=8394866;
 RA Ashkenas J., Penman M., Vasile E., Acton S., Freeman M.W.,
 RA Krieger M.;
 RA "Structures and high and low affinity ligand binding properties of
 RT murine type I and type II macrophage scavenger receptors.";
 RL J. Lipid Res. 34:983-1000(1993).
 RN [2]
 RN SEQUENCE FROM N.A. (SHORT FORM).
 RX MEDLINE=93131972; PubMed=8380589;
 RA Dol T., Wada Y., Kodama T., Higashi K.I., Kurihara Y.,
 RA Miyazaki T., Nakamura H., Uesugi S., Imanishi T., Kawabe Y.,
 RA Itakura H., Yarakaki Y., Matsumoto A.;
 RT "Charged collagen structure mediates the recognition of negatively
 RT charged macromolecules by macrophage scavenger receptors.";
 RL J. Biol. Chem. 268:2126-2133(1993).
 RN [3]
 RN SEQUENCE OF 348-458 FROM N.A.
 RX MEDLINE=91062370; PubMed=1978939;
 RA Freeman M., Ashkenas J., Rees D.J., Kingsley D.M., Copeland N.G.,
 RA Jenkins N.A., Krieger M.;
 RT "An ancient, highly conserved family of cysteine-rich protein domains
 RT revealed by cloning type I and type II murine macrophage scavenger
 RT receptors.";
 RL Proc. Natl. Acad. Sci. U.S.A. 87:8810-8814(1990).
 RN [4]
 RN SEQUENCE OF 1-4 FROM N.A.
 RX MEDLINE=95393388; PubMed=7666008;
 RA Afting R.P., Freeman M.W.;
 RT "Structure of the murine macrophage scavenger receptor gene and
 RT evaluation of sequences that regulate expression in the macrophage
 RT cell line, P388D.";
 RL J. Lipid Res. 36:1305-1314(1995).
 CC -1- FUNCTION: MEMBRANE GLYCOPROTEINS IMPLICATED IN THE PATHOLOGIC
 CC DEPOSITION OF CHOLESTEROL IN ARTERIAL WALLS DURING ATHEROGENESIS.
 CC TWO TYPES OF RECEPTOR SUBUNITS EXIST. THESE RECEPTORS MEDIATE THE
 CC ENDOCYTOSIS OF A DIVERSE GROUP OF MACROMOLECULES, INCLUDING
 CC MODIFIED LOW DENSITY LIPOPROTEINS (LDL).
 CC -1- SUBUNIT: HOMOTRIMER.
 CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN.
 CC -1- ALTERNATIVE PRODUCTS: THE TWO FORMS OF MACROPHAGE SCAVENGER
 CC RECEPTOR (TYPES I AND II) ARE PRODUCED BY ALTERNATIVE SPLICING
 CC OF THE SAME GENE.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: L04274: AAA39747.1: ALT_INIT.
 DR EMBL: L04275: AAA39748.1: ALT_INIT.
 DR EMBL: D13382: BAA02650.1: -
 DR EMBL: M59445: AAA37464.1: -
 DR EMBL: M59446: AAA37465.1: -
 DR EMBL: U13873: AAC13774.1: -
 DR PIR: A38260: A38260.
 DR MGD: MGI:98257: Scvr.
 DR InterPro: IPR000087: -
 DR InterPro: IPR001190: -

DR Pfam: PF01391; Collagen; 1.
 DR Pfam: PF00530; SRCR; 1.
 DR PRINTS: PR00258; SPERACTRCPTR.
 DR PROSITE: PS00420; SRCR_1; 1.
 KW Transmembrane; Glycoprotein; Endocytosis; Coiled coil; LDL;
 KW Heptad repeat pattern; Receptor; Alternative splicing.
 FT DOMAIN 1 55
 FT TRANSMEM 56 78
 FT DOMAIN 79 458
 FT DOMAIN 79 114
 FT DOMAIN 114 276
 FT DOMAIN 277 350
 FT DOMAIN 351 458
 FT CARBOHYD 94 94
 FT CARBOHYD 107 107
 FT CARBOHYD 147 147
 FT CARBOHYD 188 188
 FT CARBOHYD 253 253
 FT CARBOHYD 271 271
 FT VARSPPLIC 352 354
 FT VARSPPLIC 355 458
 SO SEQUENCE 458 AA; 50130 MW; 6D4C041C27EE50B CRC64;
 Query Match 60.3%; Score 55.5; DB 1; Length 458;
 Best Local Similarity 50.0%; Pred. No. 0.4;
 Matches 12; Conservative 1; Mismatches 2; Indels 9; Gaps 1;
 QY 1 GWKGD-----KGNPGRPAP 15
 Db 307 GVKGDRGQIGFPGGRCNPGAP 330

Search completed: May 23, 2001, 16:17:17
 Job time: 83 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 23, 2001, 16:15:29 ; Search time 17.92 seconds
(without alignments)
98.109 Million cell updates/sec

Title: US-09-529-691-1

Perfect score: 92
Sequence: 1 GVKGDKNGMPCGAP 15

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 374700 seqs, 117207915 residues

Total number of hits satisfying chosen parameters: 374700

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :
1: SP:archaea:*
2: SP:bacteria:*
3: SP:fungi:*
4: SP:human:*
5: SP_invertebrate:*
6: SP_mammal:*
7: SP_mhc:*
8: SP_organelle:*
9: SP_phage:*
10: SP_plant:*
11: SP_rodent:*
12: SP_unclassified:*
13: SP_vertebrate:*
14: SP_virus:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	69	75.0	886	4 Q9NUB7	Q9NUB7 homo sapien
2	64	67.3	4 Q14052	Q14052 homo sapien	Q14052 homo sapien
3	63	68.5	174	13 Q90585	Q90585 gallus gall
4	63	68.5	979	13 Q919K3	Q919K3 gallus gall
5	63	68.5	2912	4 Q14054	Q14054 homo sapien
6	62	67.4	290	5 Q26054	Q26054 paracentrot
7	62	67.4	1414	5 Q26634	Q26634 strongyloce
8	62	67.4	1802	5 Q17163	Q17163 brugia mala
9	61	66.3	224	5 Q9VYV2	Q9VYV2 drosophila
10	61	66.3	447	4 Q16593	Q16593 homo sapien
11	60	65.2	310	13 Q90612	Q90612 gallus gall
12	60	65.2	886	13 Q92029	Q92029 gallus gall
13	60	65.2	1497	4 Q9UMD9	Q9UMD9 homo sapien
14	60	65.2	1497	4 Q9NCK9	Q9NCK9 homo sapien
15	60	65.2	1532	4 Q02802	Q02802 homo sapien
16	60	65.2	1761	5 Q18407	Q18407 drosophila
17	60	65.2	1940	5 Q9VNV5	Q9VNV5 drosophila
18	60	65.2	2944	11 Q63870	Q63870 mus musculu
19	59.5	64.7	907	5 Q26312	Q26312 strongyloce

20	59.5	64.7	1752	5 Q07265	Q07265 strongyloce
21	59	64.1	771	4 Q9UJC7	Q9UJC7 homo sapien
22	59	64.1	1431	11 Q9JMH4	Q9JMH4 mesocricetu
23	59	64.1	1433	11 Q07563	Q07563 mus musculu
24	59	64.1	1682	11 Q902R9	Q902R9 mus musculu
25	59	64.1	1737	11 Q9J104	Q9J104 rattus norv
26	58	63.0	464	13 Q90412	Q90412 brachydania
27	58	63.0	1336	4 Q9Y607	Q9Y607 homo sapien
28	58	63.0	1516	4 Q9Y608	Q9Y608 homo sapien
29	58	63.0	1758	5 Q19098	Q19098 caenorhabdl
30	58	63.0	1759	5 Q19099	Q19099 caenorhabdl
31	57	62.0	132	4 P78429	P78429 homo sapien
32	57	62.0	446	4 Q9NP19	Q9NP19 homo sapien
33	57	62.0	483	11 Q9WUB9	Q9WUB9 mesocricetu
34	57	62.0	650	5 Q17866	Q17866 caenorhabdl
35	57	62.0	684	5 P90679	P90679 arenicola m
36	57	62.0	778	5 Q9U9K5	Q9U9K5 caenorhabdl
37	57	62.0	890	5 Q77087	Q77087 alvinella p
38	57	62.0	940	13 Q93484	Q93484 oncorhynch
39	57	62.0	1117	5 Q9U9K7	Q9U9K7 caenorhabdl
40	57	62.0	1549	11 Q60444	Q60444 cricetus
41	57	62.0	1669	11 Q902S0	Q902S0 mus musculu
42	57	62.0	1739	11 Q9J1I2	Q9J1I2 mus musculu
43	57	62.0	1745	4 Q9N206	Q9N206 homo sapien
44	56	60.9	284	4 Q9UC14	Q9UC14 homo sapien
45	56	60.9	295	11 Q9Z1K4	Q9Z1K4 rattus norv

ALIGNMENTS

RESULT 1
Q9NUB7 PRELIMINARY; PRT; 886 AA.
AC Q9NUB7;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE DA24A23.1 (COLLAGEN, TYPE IV, ALPHA 5 (ALPORT SYNDROME)) (FRAGMENT).
GN COL4A5.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Codley V.;
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL035425; CAB90289.1; -
KW collagen.
FT NON_TER
SQ SEQUENCE 886 AA: 85478 MW: 8C06B9FCA9AA6569 CRC64;

Query Match 75.0%; Score 69; DB 4; Length 886;
Best Local Similarity 73.3%; Pred. No. 0.023;
Matches 11; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
QY 1 GVKGDKNGMPCGAP 15
Db 480 GIKGDKNGMPCGAP 494
RESULT 2
Q14052 PRELIMINARY; PRT; 673 AA.
AC Q14052;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE ALPHA-2 TYPE IV COLLAGEN (FRAGMENT).
GN COL4A2.
OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=PLACENTA;
 RA MEDLINE=88085168; PubMed=3692475;
 RT Killen P.D., Francomano C.A., Yamada Y., Modi W.S., O'Brien S.J.;
 RT "Partial structure of the human alpha 2(IV) collagen chain and
 RT chromosomal localization of the gene (COL4A2).";
 RL Hum. Genet. 77:318-324(1987).
 DR EMBL: M24766; AAA52043.1; -;
 DR INTERPRO: IPR000087; -;
 DR INTERPRO: IPR001442; -;
 DR PFAM: PF01391; Collagen; 7.
 DR PFAM: PF01413; CA; 2.
 FT NON_TER 1
 SQ SEQUENCE 673 AA; 67174 MW; D2F3CB3111A3105 CRC64;

Query Match 69.6%; Score 64; DB 4; Length 673;
 Best Local Similarity 73.3%; Pred. No. 0.099;
 Matches 11; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 1 GYKGDGNGPWGAP 15
 DB 329 GPKGPKGDPGPGAP 343

RESULT 3
 OY0585 PRELIMINARY; PRT; 174 AA.
 AC 090585;
 DT 01-NOV-1996 (TREMBlrel. 01, Created)
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
 DT 01-JUN-2000 (TREMBlrel. 14, Last annotation update)
 DE CARILAGE ALPHA-1(IX) COLLAGEN-PROTEOGLYCAN (FRAGMENT).
 OS Gallus gallus (Chicken).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
 OC Gallus.
 OX NCBI_TaxID=9031;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA MEDLINE=90062114; PubMed=2584206;
 RA Nishimura I., Muragaki Y., Olsen B.R.;
 RT "Tissue-specific forms of type IX collagen-proteoglycan arise from the
 RT use of two widely separated promoters.";
 RL J. Biol. Chem. 264:20033-20041(1989).
 DR EMBL: M28650; AAA8709.1; -;
 DR INTERPRO: IPR000087; -;
 DR PFAM: PF01391; Collagen; 2.
 KW Extracellular matrix.
 FT NON_TER 174
 SQ SEQUENCE 174 AA; 16240 MW; C0E64A0748A224F0 CRC64;

Query Match 68.5%; Score 63; DB 13; Length 174;
 Best Local Similarity 66.7%; Pred. No. 0.034;
 Matches 10; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

OY 1 GYKGDGNGPWGAP 15
 DB 46 GIDGDKSGPAGPSP 60

RESULT 4
 OY19K3 PRELIMINARY; PRT; 979 AA.
 AC 0919K3;
 DT 01-OCT-2000 (TREMBlrel. 15, Created)
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
 DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)

DE COLLAGEN IV A1 CHAIN (FRAGMENT).
 OS Gallus gallus (Chicken).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
 OC Gallus.
 OX NCBI_TaxID=9031;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Halter W.M., Dong S.;
 RT "Composition, synthesis and assembly of the embryonic chick retinal
 RT basal lamina.";
 RL Dev. Biol. 0:0-0(2000).
 DR EMBL: AF239838; AAF44681.1; -;
 FT NON_TER 1
 SQ SEQUENCE 979 AA; 95020 MW; 5B1017D911ED4299 CRC64;

Query Match 68.5%; Score 63; DB 13; Length 979;
 Best Local Similarity 60.0%; Pred. No. 0.21;
 Matches 9; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

OY 1 GYKGDGNGPWGAP 15
 DB 573 GAKGEGSDGWPGR 587

RESULT 5
 OY1054 PRELIMINARY; PRT; 2912 AA.
 AC 014054;
 DT 01-NOV-1996 (TREMBlrel. 01, Created)
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
 DT 01-JUN-2000 (TREMBlrel. 14, Last annotation update)
 DE COLLAGEN TYPE VII PRECURSOR.
 GN COL7A1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=PLACENTA;
 RA MEDLINE=9134380; PubMed=1871109;
 RA Parente M.G., Chung L.C., Rynanen J., Woodley D.T., Wynn K.W.,
 RA Bauer E.A., Mattei M.G., Chu M.L., Uitto J.;
 RT "Human type VII collagen: cDNA cloning and chromosomal mapping of the
 RT gene.";
 RL Proc. Natl. Acad. Sci. U.S.A. 88:6931-6935(1991).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=PLACENTA;
 RA MEDLINE=9338437; PubMed=1307247;
 RA Cristiano A.M., Rosenbaum L.M., Chung-Honet L.C., Parente M.G.,
 RA Woodley D.T., Pan T.C., Zhang R.Z., Chu M.L., Burgess R.E., Uitto J.;
 RT "The large non-collagenous domain (NC-1) of type VII collagen is
 RT amino-terminal and chimeric. Homology to cartilage matrix protein, the
 RT type III domains of fibronectin and the A domains of von Willebrand
 RT factor.";
 RL Hum. Mol. Genet. 1:475-481(1992).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC TISSUE=PLACENTA;
 RA MEDLINE=93271985; PubMed=8499916;
 RA Greenspan D.S.;
 RT "The carboxyl-terminal half of type VII collagen, including the non-
 RT collagenous NC-2 domain and intron/exon organization of the
 RT corresponding region of the COL7A1 gene.";
 RL Hum. Mol. Genet. 2:273-278(1993).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC TISSUE=PLACENTA;
 RA MEDLINE=94012661; PubMed=8407948;
 RA Hoffman G.G., Lee S., Cristiano A.M., Chung-Honet L.C., Cheng W.,

Query Match	67.4%	Score 62	DB 5	Length 290
Best Local Similarity	78.6%	Pred. NO.	0.083	
Matches 11	Conservative	1	Mismatches	2
			Indels	0
			Gaps	0

RESULT	8		
017163			
ID	017163	PRELIMINARY:	PRT: 1802 AA.
AC	Q17163;		
DT	01-NOV-1996 (TrEMBLrel. 01, Created)		
DT	01-NOV-1996 (TrEMBLrel. 01, last sequence update)		
DT	01-JUN-2000 (TrEMBLrel. 14, last annotation update)		
DE	A2 (IV) BASEMENT MEMBRANE COLLAGEN (IV).		
OS	Brugia malayi.		
OC	Eukaryota; Metazoa; Nematoda; Chromadorea; Spirurida; Filarioidae;		
OC	Onchocercidae; Brugia.		
OX	NCBI_TaxID=6279;		
BN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=APERIODIC;		
RA	MEDLINE=95364846; PubMed=7637709;		
RA	Caluagi V.R., Rajan T.V.;		
RT	"The structural organization of an alpha 2 (type IV) basement membrane		
RL	collagen gene from the filarial nematode Brugia malayi.";		
DR	Mol. Biochem. Parasitol. 70:227-229(1995).		
DR	EMBL; U07224; AAC6611.1; -.		
DR	INTERPRO; IPR000087; -.		
DR	INTERPRO; IPR001442; -.		
DR	PFAM; PF01391; Collagen; 23.		
DR	PFAM; PF01413; C4; 2.		
DR	PRODOM; PD003923; -; 2.		
QO	SEQUENCE 1802 AA; 172402 MW; 595F16554CBE2D24 CRC64;		

Query Match 67.4%; Score 62; DB 5; Length 1802;
 Best Local Similarity 71.4%; Pred. No. 0.55;
 Matches 10; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 GVGKDGKNGPMPGA 14
 I::IIII::IIII
 DB 1196 GIRGDKGSPGIPGA 1209

RESULT 9

Q9VYV2 PRELIMINARY; PRT; 224 AA.
 ID Q9VYV2
 AC Q9VYV2
 DT 01-MAY-2000 (TREMBLrel. 13, Created)
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
 DT 01-JUN-2000 (TREMBLrel. 14, Last annotation update)
 DE CG14089 PROTEIN.
 GN CG14089.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OC NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BERKELEY;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celisner S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amatiates P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Mortan J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazet R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Milos G.L.G.,
 RA Abiri J.F., Aghayani A., An H.-J., Andrews-Frankoch C., Baldwin D.,
 RA Bailett R.M., Basu P.V., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brothier P.,
 RA Burtis K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahike C., Davenport L.B., Davies P.,
 RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Foster C., Gabrielian A.E., Gary N.S., Gelbart W.M., Glasser K.,
 RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Idegami C.,
 RA Jatali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Laslo P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Mlshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
 RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reilert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Sidenkiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spredling A.C., Stapleton M., Strong R., Sun E.,
 RA Svrtkac R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster.";
 RL Science 287:2185-2195(2000).
 DR EMBL: AE003517; AAF49173.1; -;
 DR FLYBASE: FBgn0036861; CG14089.
 DR INTERPRO: IPR000087; -;
 DR INTERPRO: IPR002065; -;
 DR PFAM: PF01391; Collagen.1.
 DR PRINTS: PR01217; PRICHXTENSN.
 DR SEQUENCE 224 AA; 24308 MW; F84B9C912D8EC1CD CRC64;

Query Match 66.3%; Score 61; DB 5; Length 224;
 Best Local Similarity 66.7%; Pred. No. 0.089;
 Matches 10; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 GVGKDGKNGPMPGAP 15
 I::IIII::IIII
 DB 118 GPRGDKGHPMPGIP 132

RESULT 10

Q16593 PRELIMINARY; PRT; 447 AA.
 ID Q16593
 AC Q16593
 DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
 DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
 DE COLLAGEN-LIKE PROTEIN (447 AA) (FRAGMENT).
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OC NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=PLACENTA;
 RA Kimura S.;
 RL Submitted (FEB-1990) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=PLACENTA;
 RA Kimura S.;
 RL Submitted (MAY-1992) to the EMBL/GenBank/DBJ databases.
 DR EMBL: X14963; CA33085.1; -;
 DR EMBL: X15038; CA33142.1; -;
 DR INTERPRO: IPR000087; -;
 DR PFAM: PF01391; Collagen; 5.
 FT NON_TER 1 447
 FT NON_TER 1 447
 SO SEQUENCE 447 AA; 41829 MW; FDB207023D87CC94 CRC64;

Query Match 66.3%; Score 61; DB 4; Length 447;
 Best Local Similarity 73.3%; Pred. No. 0.18;
 Matches 11; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 GVGKDGKNGPMPGAP 15
 I::IIII::IIII
 DB 280 GOKGDAGNPGDPPT 294

RESULT 11

Q90612 PRELIMINARY; PRT; 310 AA.
 ID Q90612
 AC Q90612
 DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
 DT 01-JUN-2000 (TREMBLrel. 14, Last annotation update)
 DE HYPOTHETICAL PROTEIN (FRAGMENT).
 GN COL3A1.
 OS Gallus gallus (Chicken).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
 OC Gallus.
 OC NCBI_TaxID=9031;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC MEDLINE=94266842; PubMed=8206952;
 RA Nah H.D., Niu Z., Adams S.L.;
 RT "An alternative transcript of the chick type III collagen gene that
 does not encode type III collagen.";
 RL J. Biol. Chem. 269:16443-16448(1994).
 DR EMBL: U07974; AAA83409.1; -;

DR INTERPRO: IPR000087; -
DR PFAM: PF01391; Collagen; 5.
KW Hypothetical protein.
RT NON_TER 310 310
SQ SEQUENCE 310 AA; 27601 MW; 5C60B4360832814C CRC64;

Query Match 65.2%; Score 60; DB 13; Length 310;
Best Local Similarity 71.4%; Pred. No. 0.18;
Matches 10; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

OY 1 GVGDKGNPGMPCA 14
| | | | | | | | | | | | | | | | | |
DB 281 GVGKGRGSPGCPGA 294

RESULT 12
O92029 PRELIMINARY; PRT; 886 AA.
ID O92029;
AC O92029;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DE 01-JUN-2000 (TREMBLrel. 14, Last annotation update)
GN ALPHA-1 COLLAGEN TYPE III (FRAGMENT).
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=KIDNEY;
RX MEDLINE=94266842; PubMed=8206952;
RA Nab H.D., Miu Z., Adams S.L.;
RT "An alternative transcript of the chick type III collagen gene that
RT does not encode type III collagen."
RL J. Biol. Chem. 269:16443-16448(1994).
DR EMBL: U07973; AAA83407.1; -
DR INTERPRO: IPR000087; -
DR INTERPRO: IPR001007; -
DR PFAM: PF00093; vwc; 1.
DR PFAM: PF01391; Collagen; 11.
DR PROSITE: PS01208; vwc; 1.
FT NON_TER 886 886
SQ SEQUENCE 886 AA; 81916 MW; 3DE1D5ADC088EE8 CRC64;

Query Match 65.2%; Score 60; DB 13; Length 886;
Best Local Similarity 71.4%; Pred. No. 0.52;
Matches 10; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

OY 1 GVGDKGNPGMPCA 14
| | | | | | | | | | | | | | | | | |
DB 857 GVGKGRGSPGCPGA 870

RESULT 13
O90MD9 PRELIMINARY; PRT; 1497 AA.
ID O90MD9;
AC O90MD9;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DE 01-JUN-2000 (TREMBLrel. 14, Last annotation update)
DE 180 KDA BULLOUS PEMPHIGOID ANTIGEN 2/TYPE XVII COLLAGEN.
GN BRAG2/COL17A1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97164601; PubMed=9012408;

RA Catalica B., Pulkkinen L., Li K., Kuokkanen K., Rynanen M.,
RA McGrath J.A., Uitto J.;
RT "Cloning of the human type XVII collagen gene (COL17A1), and detection
RT of novel mutations in generalized atrophic benign epidermolysis
RT bullosa."
RL Am. J. Hum. Genet. 60:352-365(1997).
DR EMBL: U76604; AAB51499.1; -
DR EMBL: U76605; AAB51499.1; JOINED.
DR EMBL: U76566; AAB51499.1; JOINED.
DR EMBL: U76567; AAB51499.1; JOINED.
DR EMBL: U76568; AAB51499.1; JOINED.
DR EMBL: U76569; AAB51499.1; JOINED.
DR EMBL: U76570; AAB51499.1; JOINED.
DR EMBL: U76571; AAB51499.1; JOINED.
DR EMBL: U76572; AAB51499.1; JOINED.
DR EMBL: U76573; AAB51499.1; JOINED.
DR EMBL: U76574; AAB51499.1; JOINED.
DR EMBL: U76575; AAB51499.1; JOINED.
DR EMBL: U76576; AAB51499.1; JOINED.
DR EMBL: U76577; AAB51499.1; JOINED.
DR EMBL: U76578; AAB51499.1; JOINED.
DR EMBL: U76579; AAB51499.1; JOINED.
DR EMBL: U76580; AAB51499.1; JOINED.
DR EMBL: U76581; AAB51499.1; JOINED.
DR EMBL: U76582; AAB51499.1; JOINED.
DR EMBL: U76583; AAB51499.1; JOINED.
DR EMBL: U76584; AAB51499.1; JOINED.
DR EMBL: U76585; AAB51499.1; JOINED.
DR EMBL: U76586; AAB51499.1; JOINED.
DR EMBL: U76587; AAB51499.1; JOINED.
DR EMBL: U76588; AAB51499.1; JOINED.
DR EMBL: U76589; AAB51499.1; JOINED.
DR EMBL: U76590; AAB51499.1; JOINED.
DR EMBL: U76591; AAB51499.1; JOINED.
DR EMBL: U76592; AAB51499.1; JOINED.
DR EMBL: U76593; AAB51499.1; JOINED.
DR EMBL: U76594; AAB51499.1; JOINED.
DR EMBL: U76595; AAB51499.1; JOINED.
DR EMBL: U76596; AAB51499.1; JOINED.
DR EMBL: U76597; AAB51499.1; JOINED.
DR EMBL: U76598; AAB51499.1; JOINED.
DR EMBL: U76599; AAB51499.1; JOINED.
DR EMBL: U76600; AAB51499.1; JOINED.
DR EMBL: U76601; AAB51499.1; JOINED.
DR EMBL: U76602; AAB51499.1; JOINED.
DR EMBL: U76603; AAB51499.1; JOINED.
DR INTERPRO: IPR000087; -
DR PFAM: PF01391; Collagen; 6.
KW Collagen.
SQ SEQUENCE 1497 AA; 150459 MW; 9E52E1076CDA6B5C CRC64;

Query Match 65.2%; Score 60; DB 4; Length 1497;
Best Local Similarity 78.6%; Pred. No. 0.9;
Matches 11; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 GVGDKGNPGMPCA 14
| | | | | | | | | | | | | | | | | |
DB 960 GPKDKGDPGVPCA 973

RESULT 14
O9NOK9 PRELIMINARY; PRT; 1497 AA.
ID O9NOK9;
AC O9NOK9;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DE 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
DE BA16H23.2 (COLLAGEN, TYPE XVII, ALPHA 1 (BP180)).
GN COL17A1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers J.-H.C., Blatz R.G., Champagne M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abdl J.F., Agbayanl A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,
 RA Ballew R.F., Baau A.A., Baxendale J., Bayraktaroglu T., Beasley E.M.,
 RA Beeson K.Y., Berens P.V., Bertman B.P., Bhandari D., Boshakov S.,
 RA Borokova D., Borchen M.R., Bouck J., Brokstein P., Brothier P.,
 RA Butts K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 Ra de Palos B., Delcher A., Deng Z., Mays A.D., Dew I., Dlez S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Flocker C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glaser K.,
 RA Foster A., Gong F., Gorrell J., Gu H., Guo P., Harris M.,
 RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Idegwan C.,
 RA Jatalin M., Kalush F., Kappe G.H., Ke Z., Kennison J.A., Kethum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lai Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Malshina N.V., Moharr C., Morris J., Mosherfi A.,
 RA Mount S.M., Moy M., Murphy E., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclty J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Pui Y., Reese M.G.,
 RA Reinelt K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-T., Wassarman D.A., Weinstock G.M., Weissbach J.,
 RA Williams S.M., Woodgate T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveril J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,

KW Extracellular matrix; Collagen.
SQ SEQUENCE 1752 AA; 170210 MW; 1AE5AA21569346D CRC64;

Query Match 64.7%; Score 59.5; DB 5; Length 1752;
Best Local Similarity 54.2%; Pred. No. 1.3;
Matches 13; Conservative 0; Mismatches 2; Indels 9; Gaps 1;

OY 1 GVGKD-----KGNPGWPGAP 15
DB 1177 GVGKDPGRGPEGAKGNPGLPGIP 1200

RESULT 21

ID 09JUC7 PRELIMINARY; PRT; 771 AA.

AC 09JUC7; 01-MAY-2000 (TREMBlrel. 13, Created)

DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)

DE COL4A5 (COLLAGEN ALPHA 5(IV) CHAIN PRECURSOR) (FRAGMENT).

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OX NCBI_TaxID=9606;

RP SEQUENCE FROM N.A.

RA Bird C.;

RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.

DR EMBL; AL031622; CAA20937.1; -

DR INTERPRO; IPR000087; -

DR PFAM; PF01391; Collagen; 10.

KW Collagen.

FT NON_TER 1 771

SQ SEQUENCE 771 AA; 72760 MW; 9CFGBD373E545A61 CRC64;

Query Match 64.1%; Score 59; DB 4; Length 771;
Best Local Similarity 73.3%; Pred. No. 0.64;
Matches 11; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 GVGKDGKNGPWPGAP 15
DB 528 GVGKDGKELGSPGAP 542

RESULT 22

ID 09JMH4 PRELIMINARY; PRT; 1431 AA.

AC 09JMH4; 01-OCT-2000 (TREMBlrel. 15, Created)

DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)

DE COLLAGEN TYPE XVII.

OS Mesocricetus auratus (Golden hamster).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OX NCBI_TaxID=10036;

RP SEQUENCE FROM N.A.

RA Yamamoto K., Inoue N., Fujimori A., Saito T., Shinkai H., Sakiyama H.;

RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.

DR EMBL; AB027759; BAA94381.1; -

SQ SEQUENCE 1431 AA; 144579 MW; 4315631FEB2C9A5C CRC64;

Query Match 64.1%; Score 59; DB 11; Length 1431;
Best Local Similarity 66.7%; Pred. No. 1.2;
Matches 10; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

OY 1 GVGKDGKNGPWPGAP 15
DB 945 GVGKDGDPGVPGCTP 959

RESULT 23

ID 007563 PRELIMINARY; PRT; 1433 AA.

AC 007563; 01-NOV-1996 (TREMBlrel. 01, Created)

DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)

DE PROCOLLAGEN, TYPE XVII, ALPHA 1 (BULLOUS PEMPHIGOID AUTOANTIGEN

BP180) (COLLAGEN XVII).

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OX NCBI_TaxID=10090;

RP SEQUENCE FROM N.A.

RA STRAIN=BALE/C;

RL MEDLINE=93232041; PubMed=8473327;

DR L.I.K., Tamai K., Tan E.M., Uitto J.;

DR "Cloning of type XVII collagen. Complementary and genomic DNA

sequences of mouse 180-kilodalton bullous pemphigoid antigen (BPAG2)

and unusual features in the 5'-end of the gene and the 3'-untranslated

region of the mRNA."

CC -1- FUNCTION: THE COLLAGEN DOMAINS OF BP180 ALLOW THE HEMIDESMOSOME TO

FORM STABLE INTERACTIONS WITH THE CONSTITUENTS OF THE EXTRACELLULAR

MATRIX OF THE CUTANEOUS BASEMENT MEMBRANE ZONE.

CC -1- SUBCELLULAR LOCATION: A COMPONENT OF THE HEMIDESMOSOME.

CC -1- DISEASE: UPON DISRUPTION OF BP180 BY AUTOANTIBODIES THE

BULLOUS PEMPHIGOID (BP).

DR EMBL; L08407; AAA37443.1; -

DR MGI; 88450; Coll17a1.

DR INTERPRO; IPR000087; -

DR PFAM; PF01391; Collagen; 5.

KW Antigen; Cell adhesion.

SQ SEQUENCE 1433 AA; 144087 MW; B8F808832A19922C CRC64;

Query Match 64.1%; Score 59; DB 11; Length 1433;
Best Local Similarity 66.7%; Pred. No. 1.2;
Matches 10; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

OY 1 GVGKDGKNGPWPGAP 15
DB 948 GVGKDGDPGVPGCTP 962

RESULT 24

ID 09QZR9 PRELIMINARY; PRT; 1682 AA.

AC 09QZR9; 01-MAY-2000 (TREMBlrel. 13, Created)

DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)

DE ALPHA 4 COLLAGEN IV.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OX NCBI_TaxID=10090;

RP SEQUENCE FROM N.A.

RA TISSUE=KIDNEY;

RC MEDLINE=20005934; PubMed=10534397;

RX Lu W., Phillips C.L., Killen P.D., Hlaing T., Harrison W.R.;

RA Elder F.F.B., Miner J.H., Overbeek P.A., Weisler M.H.;
 RT "Insertional mutation of the collagen genes col4a3 and col4a4 in a
 mouse model of alport syndrome.";
 RL Genomics 61:113-124(1999).
 DR EMBL: AF169388; AAD50450.1; -;
 DR MGD: MG1:104687; Col4a4.
 DR INTERPRO: IPR000087; -;
 DR INTERPRO: IPR001442; -;
 DR PFAM: PF01413; C4; 2.
 KW Collagen.
 SQ .SEQUENCE 1682 AA; 164096 MW; 6F7B679EDD76E904 CRC64;

Query Match

Best Local Similarity 64.1%; Score 59; DB 11; Length 1682;
 Best Local Similarity 60.0%; Pred. No. 1.4;

Matches 9; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 GYKGDKNPGMPCAP 15
 1::11:11111
 DB 176 G10GDRGDRPGPLP 190

RESULT 25

09J104 PRELIMINARY; PRT; 1737 AA.
 AC 09J104;
 DT 01-OCT-2000 (Tremblrel. 15; Created)
 DT 01-OCT-2000 (Tremblrel. 15; Last sequence update)
 DT 01-OCT-2000 (Tremblrel. 15; Last annotation update)
 DE ALPHA 4 TYPE V COLLAGEN.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID-10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=SPRAGUE DAWLEY;
 RA Chennousov M.A., Rothblum K.N., Tyler W.A., Stahl R.C., Carey D.J.;
 RT "Schwann cells synthesize type V collagen that contains a novel alpha
 4 chain.";
 RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF272661; AAF76432.1; -;
 KW Collagen.
 SQ SEQUENCE 1737 AA; 171574 MW; D635D5D57481C257 CRC64;

Query Match

Best Local Similarity 64.1%; Score 59; DB 11; Length 1737;
 Best Local Similarity 76.9%; Pred. No. 1.5;

Matches 10; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 GYKGDKNPGMPCAP 13
 1::11:11111
 DB 740 GPKGDKNPGPLP 752

Search completed: May 23, 2001, 16:17:05
 Job time: 96 sec

THIS PAGE BLANK (USPTO)